

Thu May 1 15:54:21 2003

Return

us-09-446-543a-73.rag

SEQ ID NO: 73
Database: A-Geneseg-101002
AC NO: AAW31394
File City
Page 1
Printed on
paper #16

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 30.5938 Seconds
(without alignments)
95.821 Million cell updates/sec

Title: US-09-446-543a-73
Perfect score: 109
Sequence: 1 TPDINPAMWYXXRGIRPVGRFX 22

Scoring table: BLOSUM62
Gapop 10.0, Gapept 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

- Database :
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
 - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	96.3	20	AAW31394	Human type G prote
2	105	96.3	20	AAW97236	Human type G prote
3	105	96.3	20	AAW10365	Human type G prote
4	105	96.3	20	AAW49294	Human oxytocin sec
5	105	96.3	20	AAW62534	19P2 ligand peptid
6	105	96.3	20	AAW90992	Human CRH releasin
7	105	96.3	21	AAW31395	Proactin releasin
8	105	96.3	21	AAW10366	Human type G prote
9	105	96.3	21	AAW62535	Human oxytocin sec
10	105	96.3	22	AAW31396	Human CRH releasin
					Human type G prote

11	105	96.3	22	AAW10367	Human oxytocin sec
12	105	96.3	22	AAW62536	Human CRH releasin
13	105	96.3	31	AAW31391	Human type G prote
14	105	96.3	31	AAW97235	Human type G prote
15	105	96.3	31	AAW87615	Human type G prote
16	105	96.3	31	AAW10362	Human type G prote
17	105	96.3	31	AAW49291	Human oxytocin sec
18	105	96.3	31	AAW62531	19P2 ligand peptid
19	105	96.3	31	AAW90991	Human CRH releasin
20	105	96.3	31	AAW90995	Proactin releasin
21	105	96.3	32	AAW31392	Human type G prote
22	105	96.3	32	AAW10363	Human oxytocin sec
23	105	96.3	32	AAW62532	Human CRH releasin
24	105	96.3	33	AAW31393	Human type G prote
25	105	96.3	33	AAW10364	Human oxytocin sec
26	105	96.3	33	AAW62533	Human CRH releasin
27	105	96.3	37	AAW31390	Human CRH releasin
28	105	96.3	87	AAW97226	Human type G prote
29	105	96.3	87	AAW10361	Human oxytocin sec
30	105	96.3	87	AAW62530	Human CRH releasin
31	104	95.4	20	AAW31387	Human type G prote
32	104	95.4	20	AAW31374	Rat type G prote
33	104	95.4	20	AAW97232	Bovine G prote
34	104	95.4	20	AAW97234	Bovine pituitary-d
35	104	95.4	20	AAW95191	Rat type G prote
36	104	95.4	20	AAW95175	Bovine pituitary-d
37	104	95.4	20	AAW10350	Murine pituitary-d
38	104	95.4	20	AAW10358	Bovine oxytocin se
39	104	95.4	20	AAW49301	Rat oxytocin secre
40	104	95.4	20	AAW49302	19P2 ligand peptid
41	104	95.4	20	AAW62519	Bovine CRH releasi
42	104	95.4	20	AAW62527	Rat CRH releasin
43	104	95.4	20	AAW90994	Proactin releasin
44	104	95.4	20	AAW90996	Proactin releasin
45	104	95.4	20	AAW46954	Peptide Prnp20 fra

ALIGNMENTS

RESULT 1
AAW31394
ID AAW31394 standard; Peptide: 20 AA.
XX
AC AAW31394;
XX
XX 06-APR-1998 (first entry)
DT
DE Human type G protein-coupled receptor ligand fragment 4.
XX
XX G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.
XX
XX Homo sapiens.
XX
XX W09724436-A2.
XX
XX 10-JUL-1997.
XX
XX 26-DEC-1996; 96W0-JP03821.
XX
XX 18-SEP-1996; 96JP-0246573.
XX 28-DEC-1995; 95JP-0343371.
XX 15-MAR-1996; 96JP-0059419.
XX 12-AUG-1996; 96JP-0211805.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
XX Kawamata Y, Kikada C;
XX WPI; 1997-363672/33.

DR N-PSDB; AAV02431.

XX Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 XX
 PS Claim 2; Page 185; 258pp; English.

CC This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 34 to 53 of the
 CC sequence represented in AAV01390 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC propylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spino-cerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosacchara. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.

SQ Sequence 20 AA;

Query Match 96.3%; Score 105; DB 18; Length 20;
 Best Local Similarity 90.0%; Pred. No. 3.2e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMWYXXRGIRPVGRF 20
 ||||| |||||
 DB 1 TPDINPAMWYASRGIRPVGRF 20

RESULT 2

AAW97236

ID AAW97236 standard; peptide; 20 AA.

AC AAW97236;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX Rat type ligand; modulation; prolactin secretion;
 KW G protein-coupled receptor; GPCR; hypocoovarianism; lactation;
 KW menopausal syndrome; euthyroid; hypometabolism; lactation;
 KW pituitary adenomatosis; brain tumour; emmenopathy; autolimmune disease;
 KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
 KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
 KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
 KW contraceptive; placental function; chorioncarinoma; hydatid mole;
 KW irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
 KW abnormal lipidmetabolism; oxytocia.

OS Homo sapiens.

PN W09858962-A1.

PD 30-DEC-1998.

PE 22-JUN-1998; 98WO-JP02765.

PR 23-JUN-1997; 97JP-0165437.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;

XX

DR WPI; 1999-105614/09.

XX Use of G protein-coupled receptor ligands - for modulating prolactin
 PT secretion or placental function, e.g. for treating menopausal
 PT syndrome, tumours, autoimmune disease or abnormal pregnancy
 XX
 PS Claim 3; Page 166; 241pp; English.

CC The present sequence represents a human type ligand fragment. It
 CC is used in the course of the invention. The specification describes
 CC an agent for modulating prolactin secretion which comprises a
 CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
 CC protein. The agents for promoting prolactin secretion can be used for
 CC treating or preventing hypocoovarianism, gonocyst cacosgenesis, menopausal
 CC syndrome, euthyroid or hypometabolism. They can be used for promoting
 CC lactation in a domestic mammal and as an aphrodisiac. The agents for
 CC inhibiting prolactin secretion can be used for treating or preventing
 CC pituitary adenomatosis, brain tumour, emmenopathy, autolimmune disease,
 CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
 CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
 CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
 CC The inhibitory agents can also be used as contraceptives. The agents for
 CC modulating placental function can be used for treating or preventing
 CC chorioncarinoma, hydatid mole, irruption mole, abortion, unthrifty fetus,
 CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.

SQ Sequence 20 AA;

Query Match 96.3%; Score 105; DB 20; Length 20;
 Best Local Similarity 90.0%; Pred. No. 3.2e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMWYXXRGIRPVGRF 20
 ||||| |||||
 DB 1 TPDINPAMWYASRGIRPVGRF 20

RESULT 3

AAB10365

ID AAB10365 standard; peptide; 20 AA.

AC AAB10365;

DT 24-NOV-2000 (first entry)

DE Human oxytocin secretion promoting peptide SEQ ID NO: 35.

XX Human; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.

OS Homo sapiens.

PN W0200038704-A1.

PD 06-JUL-2000.

PE 22-DEC-1999; 99WO-JP07199.

PR 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-452298/39.

XX Physiologically active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine

PS Disclosure; Page 63; 72pp; Japanese.
XX
XX This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.,
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a human peptide which acts as an oxytocin secretion
CC promoter.
XX
SQ Sequence 20 AA;
Query Match 96.3%; Score 105; DB 21; Length 20;
Best Local Similarity 90.0%; Pred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TPDINPAMYXXRGIRPVGRF 20
Db 1 TPDINPAMYASRGIRPVGRF 20
|||||
RESULT 4:
ID AAY49294 standard; peptide; 20 AA.
XX
AC AAY49294;
XX
XX 22-FEB-2000 (first entry)
DE 19P2 ligand peptide fragment.
XX
XX Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KM pituitary; regulatory mechanism; central nervous system; pancreatic.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Modified-site 20
XX /note="C-terminal amide"
XX
PN WO960112-A1.
XX
XX 25-NOV-1999.
PD
XX
XX 20-MAY-1999; 99WO-JP02650.
PF
XX
XX 21-MAY-1998; 98JP-0140293.
PR
XX
XX (TAKE) TAKEDA CHEM IND LTD.
PA
XX
XX Matsumoto H, Kitada C, Hinuma S;
PI
XX
XX WPI; 2000-039381/03.
DR
XX
XX New monoclonal antibodies, useful in diagnosis, as drugs and in
PT studying diseases related to ligand abnormality
XX
XX
PS Disclosure; Page 26; 73pp; Japanese.
XX
XX The invention provides a monoclonal antibody which has a specific
CC reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC derivative. The antibodies can be used in diagnosis or to treat or
CC prevent diseases associated with abnormality in the pituitary function
CC regulatory mechanism (e.g. promotion of prolactin secretion), central
CC nervous regulatory mechanism, and pancreatic function regulatory
CC mechanism. The antibody-based immunoassay can also be applied in
CC clarifying the physiological functions of the ligand and its derivative.
CC Sequences AAY49290-302 represent peptide fragments of the 19P2 ligand.
XX
SQ Sequence 20 AA;

Query Match 96.3%; Score 105; DB 21; Length 20;
Best Local Similarity 90.0%; Pred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TPDINPAMYXXRGIRPVGRF 20
Db 1 TPDINPAMYASRGIRPVGRF 20
|||||
RESULT 5:
ID AAG62534 standard; peptide; 20 AA.
XX
XX AAG62534;
AC
XX
XX 24-NOV-2001 (first entry)
DT
XX
XX Human CRH releasing protein related peptide SEQ ID NO: 35.
DE
XX
XX Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;
KM analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;
KM Addison's disease; adrenal gland hyperfunction; obesity.
XX
XX Homo sapiens.
OS
XX
XX WO200135984-A1.
PN
XX
XX 25-MAY-2001.
PD
XX
XX 17-NOV-2000; 2000WO-JP08119.
PF
XX
XX 18-NOV-1999; 99JP-0327900.
PR
XX
XX 26-SEP-2000; 2000JP-0297073.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
PA
XX
XX Kitada C, Matsumoto H, Hinuma S;
PI
XX
XX WPI; 2001-35552/37.
DR
XX
XX Use of G protein receptor ligand or peptide for controlling
PT corticotrophin releasing hormone secretion
XX
XX
PS Claim 4; Page 75; 90pp; Japanese.
XX
XX The present sequence describes a method of controlling the secretion of
CC corticotrophin releasing hormone (CRH), involving the use of a G protein
CC receptor ligand. This can be used to control the secretion of CRH and is
CC useful as an analgesic or for treating, preventing or ameliorating
CC diseases associated with CRH secretion such as hyperaldosteronism,
CC hypercortisolemia, secondary or chronic hypoadrenocorticism, Addison's
CC disease (including boredom, nausea, pigmentation, hypogonadism, hair
CC loss, and hypotension), adrenal gland hypofunction and obesity. The
CC present sequence is a peptide used in the exemplification of the
XX invention.
XX
SQ Sequence 20 AA;
Query Match 96.3%; Score 105; DB 22; Length 20;
Best Local Similarity 90.0%; Pred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TPDINPAMYXXRGIRPVGRF 20
Db 1 TPDINPAMYASRGIRPVGRF 20
|||||
RESULT 6:
ID AAB90992 standard; Peptide; 20 AA.
XX
XX AAB90992;
AC

XX 22-JUN-2001 (first entry)
 XX Prolactin releasing peptide SEQ ID NO:166.
 DE Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimide; maleimide group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200069900-A2.
 XX 23-NOV-2000.
 PD 17-MAY-2000; 2000MO-US13576.
 PF 17-MAY-1999; 99US-0134406.
 PR 10-SEP-1999; 99US-0153406.
 PR 15-OCT-1999; 99US-0159783.
 XX (CONU-) CONUICHEM INC.
 PA Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudau K;
 PI WPI; 2001-112059/12.
 DR Modifying and attaching therapeutic peptides to albumin prevents
 PT peptidase degradation, useful for increasing length of in vivo activity
 PS Disclosure; Page 244; 733pp; English.
 XX The present invention describes a modified therapeutic peptide (I)
 CC comprising a therapeutically active amino acid region (III) and a
 CC reactive group (II) (e.g. succinimide and maleimide groups) attached to
 CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity
 CC in vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention.
 CC XX
 SQ Sequence 20 AA:
 Query Match 96.3%; Score 105; DB 22; Length 20;
 Best Local Similarity 90.0%; Pred. No. 3.2e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TPDINPAMYXXRGIRPVGRF 20
 ||||||||| |||||||||
 Db 1 TPDINPAMYASRGIRPVGRF 20
 ||||||||| |||||||||
 RESULT 7
 ID AAW31395 standard; Peptide: 21 AA.
 AC AAW31395;
 XX 06-APR-1998 (first entry)
 DE Human type G protein-coupled receptor ligand fragment 5.
 XX

KW G protein-coupled receptor; ligand binding; pharmaceutical;
 KW modulator; pituitary; central nervous system; pancreas; prophylactic;
 KW therapeutic agent.
 XX Homo sapiens.
 OS WO9724436-A2.
 XX 10-JUL-1997.
 PD 26-DEC-1996; 96WO-JP03821.
 PF 18-SEP-1996; 96JP-0246573.
 PR 28-DEC-1995; 95JP-0343371.
 PR 15-MAR-1996; 96JP-0059419.
 PR 12-AUG-1996; 96JP-0211805.
 XX (TAKE) TAKEEDA CHEM IND LTD.
 PA Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
 PI Kawamata Y, Kitada C;
 DR WPI; 1997-363672/33.
 DR N-PSDB; AAV02432.
 XX Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 PS Claim 2; Page 186; 258pp; English.
 XX This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 34 to 54 of the
 CC sequence represented in AAW31390 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligosaccharia. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 CC XX
 SQ Sequence 21 AA:
 Query Match 96.3%; Score 105; DB 18; Length 21;
 Best Local Similarity 90.0%; Pred. No. 3.4e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TPDINPAMYXXRGIRPVGRF 20
 ||||||||| |||||||||
 Db 1 TPDINPAMYASRGIRPVGRF 20
 ||||||||| |||||||||
 RESULT 8
 ID AAB10366 standard; peptide: 21 AA.
 AC AAB10366;
 XX 24-NOV-2000 (first entry)
 DE Human oxytocin secretion promoting peptide SEQ ID NO: 36.
 XX Human; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 XX

KM caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX Homo sapiens.
OS WO200038704-A1.
XX
XX
XX 06-JUL-2000.
XX
XX
XX 22-DEC-1999; 99WO-JP07199.
XX
XX 25-DEC-1998; 98JP-0369585.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Matsumoto H, Kitada C, Hinuma S;
XX WPI; 2000-452298/39.
XX
XX Physiologically-active polypeptide recognized as ligand by G
XX protein-coupled receptor protein, for promoting secretion of oxytocin,
XX as drugs for diseases relating to oxytocin secretion and in veterinary
XX medicine -
XX
XX
XX Disclosure; Page 63; 72pp; Japanese.
XX
XX This invention describes a novel oxytocin secretion-regulating agent
XX which contains a ligand peptide or its salt for the G protein-coupled
XX receptor protein. It is useful in the form of drugs for ameliorating,
XX preventing and treating diseases relating to oxytocin secretion e.g.,
XX weak pains and atonic bleeding, before and after expulsion of placenta,
XX uterine recovery failure, caesarean section, stoppage of artificial
XX fertilization or galactostasis and is also applicable in veterinary
XX medicine for promoting milk production in cow, goat and pig. This
XX sequence represents a human peptide which acts as an oxytocin secretion
XX promoter.
XX
XX
SQ Sequence 21 AA:

Query Match 96.3%; Score 105; DB 21; Length 21;
Best Local Similarity 90.0%; Pred. No. 3,4e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAWYXXRGIRPYGRF 20
DB 1 TPDINPAWYASRGIRPYGRF 20

RESULT 9
AAG62535
ID AAG62535 standard; peptide; 21 AA.
XX
XX AAG62535;
XX
XX 24-AUG-2001 (first entry)
XX
XX Human CRH releasing protein related peptide SEQ ID NO: 36.
XX
XX Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;
XX analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;
XX Addison's disease; adrenal gland hyperfunction; obesity.
XX
XX Homo sapiens.
XX
XX WO200135984-A1.
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-JP08119.
XX
XX 18-NOV-1999; 99JP-0327900.
XX
XX 26-SEP-2000; 2000JP-0297073.
XX

PA (TAKE) TAKEDA CHEM IND LTD.
XX
XX
XX Kitada C, Matsumoto H, Hinuma S;
XX WPI; 2001-355552/37.
XX
XX
XX Use of G protein receptor ligand or peptide for controlling
XX corticotrophin releasing hormone secretion -
XX
XX
XX Disclosure; Page 75; 90pp; Japanese.
XX
XX
XX The present sequence describes a method of controlling the secretion of
XX corticotrophin releasing hormone (CRH), involving the use of a G protein
XX receptor ligand. This can be used to control the secretion of CRH and its
XX useful as an analgesic or for treating, preventing or ameliorating
XX diseases associated with CRH secretion such as hyperaldosteronism,
XX hypercortisolemia, secondary or chronic hypoadrenocorticism, Addison's
XX disease (including boredom, nausea, pigmentation, hypogonadism, hair
XX loss, and hypotension), adrenal gland hypofunction and obesity. The
XX present sequence is a peptide used in the exemplification of the
XX invention.
XX
XX
SQ Sequence 21 AA:

Query Match 96.3%; Score 105; DB 22; Length 21;
Best Local Similarity 90.0%; Pred. No. 3,4e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAWYXXRGIRPYGRF 20
DB 1 TPDINPAWYASRGIRPYGRF 20

RESULT 10
AAW31396
ID AAW31396 standard; peptide; 22 AA.
XX
XX AAW31396;
XX
XX 06-APR-1998 (first entry)
XX
XX Human type G protein-coupled receptor ligand fragment 6.
XX
XX G protein-coupled receptor; ligand binding; pharmaceutical;
XX modulator; pituitary; central nervous system; pancreas; prophylactic;
XX therapeutic agent.
XX
XX Homo sapiens.
XX
XX WO9724436-A2.
XX
XX 10-JUL-1997.
XX
XX 26-DEC-1996; 96WO-JP03821.
XX
XX 18-SEP-1996; 96JP-0246573.
XX
XX 28-DEC-1995; 95JP-0343371.
XX
XX 15-MAR-1996; 96JP-0059419.
XX
XX 12-AUG-1996; 96JP-0211805.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
XX Kawamata Y, Kitada C;
XX WPI; 1997-363672/33.
XX
XX N-PSDB; AAW02433.
XX
XX
XX Ligand peptide for G protein-coupled receptor - acts by modulating
XX function in the central nervous system, pancreas and pituitary gland
XX
XX Claim 2; Page 186; 258pp; English.

CC This sequence represents a peptide fragment from a novel human type
CC ligand polypeptide corresponding to amino acid residues 34 to 55 of the
CC sequence represented in AAW31390 and is used in an assay to monitor
CC ligand binding to the G protein coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a
CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
CC trauma, growth hormone secretory disease, hyper- and polyphagia,
CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
CC acute myocardial infarction, infertility, splenocerebellar degeneration,
CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
CC oligosaccharide. Assays can also be developed to screen compounds which are
CC capable of altering the binding activity of the ligand affecting
CC activation of the G protein-coupled receptor protein.

SQ Sequence 22 AA;

Query Match 96.3%; Score 105; DB 18; Length 22;
Best Local Similarity 90.0%; Pred. No. 3.5e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20

DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 11

ID AAB10367 standard; peptide; 22 AA.

AC AAB10367;

DT 24-NOV-2000 (first entry)

DE Human oxytocin secretion promoting peptide SEQ ID NO: 37.

Human; oxytocin secretion promoter; G protein-coupled receptor protein;
treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
caesarean section; artificial fertilization; galactostasis; goat; pig;
veterinary medicine; milk production.

OS Homo sapiens.

PI WO200038704-A1.

PD 06-JUL-2000.

PF 22-DEC-1999; 99WO-JP07199.

PR 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-452298/39.

Physiologically-active polypeptide recognized as ligand by G
protein-coupled receptor protein, for promoting secretion of oxytocin,
as drugs for diseases relating to oxytocin secretion and in veterinary
medicine

PS Disclosure; Page 64; 72pp; Japanese.

This invention describes a novel oxytocin secretion-regulating agent
which contains a ligand peptide or its salt for the G protein-coupled
receptor protein. It is useful in the form of drugs for ameliorating,
preventing and treating diseases relating to oxytocin secretion e.g.

CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a human peptide which acts as an oxytocin secretion
CC promoter.

SQ Sequence 22 AA;

Query Match 96.3%; Score 105; DB 21; Length 22;
Best Local Similarity 90.0%; Pred. No. 3.5e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20

DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 12

ID AAG62536 standard; peptide; 22 AA.

AC AAG62536;

DT 24-AUG-2001 (first entry)

DE Human CRH releasing protein related peptide SEQ ID NO: 37.

Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;
analgesic; hyperaldosteronism; hypercortisolaemia; hypoadrenocorticism;
Addison's disease; adrenal gland hyperfunction; obesity.

OS Homo sapiens.

PI WO200135984-A1.

PD 25-MAY-2001.

PF 17-NOV-2000; 2000WO-JP08119.

PR 18-NOV-1999; 99JP-0327900.

PR 26-SEP-2000; 2000JP-0297073.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Kitada C, Matsumoto H, Hinuma S;

DR WPI; 2001-355552/37.

Use of G protein receptor ligand or peptide for controlling
corticotrophin releasing hormone secretion

PS Disclosure; Page 75; 90pp; Japanese.

The present sequence describes a method of controlling the secretion of
corticotrophin releasing hormone (CRH), involving the use of a G protein
receptor ligand. This can be used to control the secretion of CRH and is
useful as an analgesic or for treating, preventing or ameliorating
diseases associated with CRH secretion such as hyperaldosteronism,
hypercortisolaemia, secondary or chronic hypoadrenocorticism, Addison's
disease (including boredom, nausea, pigmentation, hypogonadism, hair
loss, and hypotension), adrenal gland hypofunction and obesity. The
present sequence is a peptide used in the exemplification of the
invention.

SQ Sequence 22 AA;

Query Match 96.3%; Score 105; DB 22; Length 22;
Best Local Similarity 90.0%; Pred. No. 3.5e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20

DB 1 TPDINPAMYASRGIRPVGRF 20

.Db 1 TPDINPAMWASRGIRPVGRF 20

RESULT 13
AAW31391

ID AAW31391 standard; Peptide: 31 AA.

AC AAW31391;

DT 06-APR-1998 (first entry)

DE Human type G protein-coupled receptor ligand fragment 1.

XX G protein-coupled receptor; ligand binding; pharmaceutical;
XX modulator; pituitary; central nervous system; pancreas; prophylactic;
XX therapeutic agent.

OS Homo sapiens.

PN WO9724436-A2.

PD 10-JUL-1997.

PF 26-DEC-1996; 96WO-JP03821.

PR 18-SEP-1996; 96JP-0246573.

PR 28-DEC-1995; 95JP-0343371.

PR 15-MAR-1996; 96JP-0059419.

PR 12-AUG-1996; 96JP-0211805.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;

PI Kawamata Y, Kitada C;

DR WPI: 1997-363672/33.

DR N-PSDB; AAV02428.

PS Claim 2; Page 184; 258pp; English.

XX This sequence represents a peptide fragment from a novel human type

CC ligand polypeptide corresponding to amino acid residues 23 to 53 of the

CC sequence represented in AAW31390 and is used in an assay to monitor

CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical

CC compositions containing this ligand may be used as a pituitary function

CC modulator. This ligand could have specific applications as a

CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic

CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,

CC trauma, growth hormone secretory disease, hyper- and polyphagia,

CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,

CC Turner's syndrome, diabetes, cancer, pancreatitis, renal disease,

CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,

CC acute myocardial infarction, infertility, spinocerebellar degeneration,

CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or

CC oligosacchara. Assays can also be developed to screen compounds which are

CC capable of altering the binding activity of the ligand affecting

CC activation of the G protein-coupled receptor protein.

XX Sequence 31 AA;

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 14

AAW97235

ID AAW97235 standard; peptide: 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 14

AAW97235

ID AAW97235 standard; peptide: 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 14

AAW97235

ID AAW97235 standard; peptide: 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 14

AAW97235

ID AAW97235 standard; peptide: 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 14

AAW97235

ID AAW97235 standard; peptide: 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 14

AAW97235

ID AAW97235 standard; peptide: 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 14

AAW97235

ID AAW97235 standard; peptide: 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 15

AAW87615

ID AAW87615 standard; Peptide; 31 AA.

XX AAW87615;

AC AAW87615;

DT 29-MAR-1999 (first entry)

XX Human 19P2 ligand.

DE Human 19P2 ligand.

XX 19P2 ligand; G protein coupled receptor; pituitary;

KM prolactin releasing peptide; human; dementia; breast cancer;

KW therapy.

XX Homo sapiens.

OS Homo sapiens.

XX EP887417-A2.

PN EP887417-A2.

XX 30-DEC-1998.

PD 30-DEC-1998.

XX 25-JUN-1998; 98EP-0111725.

PF 25-JUN-1998; 98EP-0111725.

XX 27-JUN-1997; 97JP-0172118.

PR 27-JUN-1997; 97JP-0172118.

XX (TAKE) TAKEDA CHEM IND LTD.

PA (TAKE) TAKEDA CHEM IND LTD.

XX Moriya T, Nishimura O, Suenaga M, Tanaka Y;

PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;

XX WPI; 1999-047884/05.

DR WPI; 1999-047884/05.

XX Producing a 19P2 pituitary G protein receptor ligand - by cleavage

PT of a fusion protein, useful for preventing and treating dementia,

PT breast cancer, renal failure and autoimmune disease

XX Claim 5; Page 35; 56pp; English.

PS Claim 5; Page 35; 56pp; English.

XX This is the amino acid sequence of the human pituitary G

CC protein-coupled receptor ligand 19P2L. A method suitable for

CC commercial high-level production of 19P2L comprises expressing

CC the ligand in host cells as a recombinant fusion protein e.g. with

CC human basic fibroblast growth factor (see AYW83796-97) that has

CC been modified to include an N-terminal cysteine residue. The

CC ligand is released from the fusion by cyanylation followed by

CC ammonolysis. 19P2L has prolactin secretion-stimulating and (at

CC high doses) prolactin secretion-inhibiting properties. It can be

CC used in the treatment and prevention of various diseases including:

CC senile dementia, cerebrovascular dementia, and dementia associated

CC with: neurological disorders (e.g. Alzheimer's disease, Parkinson's

CC disease, Pick's disease, Huntington's disease), infectious diseases

CC (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or

CC toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,

CC intoxication by drugs, metal and organic compounds), tumorigenic

CC diseases (e.g. brain tumour), traumatic diseases (e.g. chronic

CC subarachnoid haemorrhage, and other types of dementia, depression,

CC hyperactive child syndrome (microcephalopathy) and disturbance of

CC consciousness. It is also useful for prevention and treatment of

CC diseases associated with prolactin hypo and hypersecretion

CC respectively, including: hyperprolactinaemia, pituitary adenoma,

CC breast cancer, infertility, impotence and autoimmune disease

CC (hypersecretion disorders), and seminal vesicle hypoplasia,

CC osteoporosis, menopausal syndrome and renal failure (hyposecretion

CC disorders). The 19P2 polypeptide/amide is also useful as a test

CC reagent for study of the prolactin secretory function or as a

CC lactagogue in mammalian farm animals.

XX Sequence 31 AA;

SQ Sequence 31 AA;

Query Match 96.3%; Score 105; DB 20; Length 31;

Best Local Similarity 90.0%; Pred. No. 5.2e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 12 TPDINPAMYASRGIRPYGRF 31

Search completed: May 1, 2003, 14:30:50
Job time : 31.5938 secsOY 1 TPDINPAMYASRGIRPYGRF 20
||||||| |||||

File copy

WPI; 1997-363672/33.

CC This is the amino acid sequence of the rat pituitary G
CC protein-coupled receptor ligand 19p2L. A method suitable for
CC commercial high-level production of 19p2L comprises expressing
CC the ligand in host cells as a recombinant fusion protein e.g. with
CC human basic fibroblast growth factor (see A483794-95) that has
CC been modified to include an N-terminal cysteine residue. The
CC ligand is released from the fusion by cyanilation followed by
CC ammonolysis. 19p2L has prolactin secretion-stimulating and (at
CC high doses) prolactin secretion-inhibiting properties. It can be
CC used in the treatment and prevention of various diseases including:
CC senile dementia, cerebrovascular dementia, and dementia associated
CC with: generalised disorders (e.g. Alzheimer's disease, Parkinson's
CC disease, Pick's disease, Huntington's disease), infectious diseases
CC (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or
CC toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,
CC intoxication by drugs, metal and organic compounds), tumourigenic
CC diseases (e.g. brain tumour), traumatic diseases (e.g. chronic
CC subarachnoidal haemorrhage, and other types of dementia, depression,
CC hyperactive child syndrome (microencephalopathy) and disturbance of
CC consciousness. It is also useful for prevention and treatment of
CC diseases associated with prolactin hypo and hypersecretion
CC respectively, including: hyperprolactinaemia, pituitary adenoma,
CC breast cancer, infertility, impotence and autoimmune disease
CC (hypersecretion disorders), and seminal vesicle hypoplasia,
CC osteoporosis, menopausal syndrome and renal failure (hyposecretion
CC disorders). The 19p2 polypeptide/amide is also useful as a test
CC reagent for study of the prolactin secretory function or as a
CC lactagogue in mammalian farm animals.

CC Sequence 31 AA;

Query Match 91.5%; Score 43; DB 20; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.031;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
||| ||| |
Db 1 SRAHQSMETR 11

RESULT 4
AAW95173
ID AAW95173 standard; peptide; 31 AA.

AC AAW95173;

DT 10-MAR-1999 (first entry)

DE Murine pituitary-derived ligand mature polypeptide sequence.

KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal.

OS Mus sp.

PN WO9849295-A1.

PD 05-NOV-1998.

PF 27-APR-1998; 98WO-JP01923.

PR 28-APR-1997; 97JP-0109974.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Fukusumi S, Hinuma S;

DR WPI; 1999-009423/01.

XX New polypeptide ligand for orphan G protein coupled receptors - used
PT for treating disorders of central nervous system, pituitary and
PT pancreas, and for drug screening
XX Disclosure; Page 134; 206pp; English.

XX This represents the matured murine pituitary-derived ligand polypeptide
XX sequence. The polypeptide is a ligand for the G-protein coupled orphan
XX receptor designated GPR10 (human) or UHR-1 (rat). Cells transformed with
XX a vector containing the ligand polypeptide encoding DNA are used to
XX produce a recombinant ligand polypeptide. The ligand polypeptide, and its
XX fragments, modulate function of the pituitary, central nervous system,
XX pancreas and other tissues and can be used to screen for agents that
XX modulate binding of the polypeptide to the receptor; to quantify the
XX amount of receptor in a sample and to raise antibodies. They may also be
XX used therapeutically, e.g. to treat senile dementia; Alzheimer's,
XX Parkinson's or Huntington's diseases; Creutzfeld-Jakob disease; poisoning
XX by heavy metals or drugs; diabetes; schizophrenia; disorders of growth
XX hormone secretion; cancer; rheumatoid arthritis, epilepsy and many
XX others, also to improve post-operative nutritional status and as
XX vasopressor. Transgenic animals carrying the ligand polypeptide encoding
XX DNA or its mutant are used to study the function of the polypeptide-
XX expressing genes, as models of disease, for drug screening and as source
XX of cell lines. The ligand polypeptide DNA is used as a source of probes
XX and primers; to identify related sequences; in receptor-binding assays;
XX for production of Ab and antisera; in drug development; for gene therapy
XX and to develop transgenic animals.

XX Sequence 31 AA;

Query Match 91.5%; Score 43; DB 20; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.031;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
||| ||| |
Db 1 SRAHQSMETR 11

RESULT 5
AAW95174
ID AAW95174 standard; Protein; 31 AA.

AC AAW95174;

DT 10-MAR-1999 (first entry)

DE Murine pituitary-derived ligand polypeptide antigenic epitope.

KW Pituitary-derived ligand polypeptide; G-protein coupled orphan receptor;
KW GPR10; UHR-1; modulator; pituitary; central nervous system; pancreas;
KW tissue; screen; therapeutic; binding; senile dementia; ligand; murine;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; drug;
KW Creutzfeld-Jakob disease; poisoning; schizophrenia; growth hormone;
KW secretion; diabetes; cancer; rheumatoid arthritis; epilepsy; vasopressor;
KW gene therapy; transgenic animal; epitope.

OS Mus sp.

PN WO9849295-A1.

PD 05-NOV-1998.

PF 27-APR-1998; 98WO-JP01923.

PR 28-APR-1997; 97JP-0109974.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Fukusumi S, Hinuma S;

DR WPI; 1999-009423/01.

XX New polypeptide ligand for orphan G protein coupled receptors - used
 PT for treating disorders of central nervous system, pituitary and
 PT pancreas, and for drug screening
 PS Disclosure; Page 26; 206pp; English.

XX The invention relates to a murine pituitary-derived ligand polypeptide
 CC which is a ligand for the G-protein coupled orphan receptor designated
 CC GPR10 (human) or UHR-1 (rat). Cells transformed with a vector containing
 CC the ligand polypeptide encoding DNA are used to produce a recombinant
 CC ligand polypeptide. The ligand polypeptide, and its fragments, modulate
 CC function of the pituitary, central nervous system, pancreas and other
 CC tissues and can be used to screen for agents that modulate binding of
 CC the polypeptide to the receptor; to quantify the amount of receptor in a
 CC sample and to raise antibodies. They may also be used therapeutically,
 CC e.g. to treat senile dementia; Alzheimer's, Parkinson's or Huntington's
 CC diseases; Creutzfeldt-Jakob disease; poisoning by heavy metals or drugs;
 CC diabetes; schizophrenia; disorders of growth hormone secretion; cancer;
 CC rheumatoid arthritis, epilepsy and many others, also to improve post-
 CC operative nutritional status and as vasopressor. Transgenic animals
 CC carrying the ligand polypeptide encoding DNA or its mutagen are used to
 CC study the function of the polypeptide-expressing genes, as models of
 CC disease, for drug screening and as source of cell lines. The ligand
 CC polypeptide DNA is used as a source of probes and primers; to identify
 CC related sequences; in receptor-binding assays; for production of Ab and
 CC antisera; in drug development; for gene therapy and to develop
 CC transgenic animals. Sequences AAW95174 to AAW95178 represent antigenic
 CC epitopes which can be used for the preparation of anti-ligand polypeptide
 CC antibody.
 CC
 SQ Sequence 31 AA;

Query Match 91.5%; Score 43; DB 20; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
 || || || || ||
 DB 1 SRAHQHSMETR 11

RESULT 6
 AAB10355
 ID AAB10355 standard; peptide; 31 AA.

AC AAB10355;

DT 24-NOV-2000 (first entry)

DE Rat oxytocin secretion promoting peptide SEQ ID NO: 18.

XX Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.

OS Rattus sp.

PN WO2000038704-A1.

PD 06-JUL-2000.

PF 22-DEC-1999; 99WO-JP07199.

PR 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-452298/39.

PT Physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine
 PS Claim 3; Page 57; 72pp; Japanese.

XX This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a rat peptide which acts as an oxytocin secretion
 CC promoter.
 CC
 SQ Sequence 31 AA;

Query Match 91.5%; Score 43; DB 21; Length 31;
 Best Local Similarity 72.7%; Pred. No. 0.031;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
 || || || || ||
 DB 1 SRAHQHSMETR 11

RESULT 7
 AAY87504
 ID AAY87504 standard; protein; 31 AA.

AC AAY87504;

DT 18-JUL-2000 (first entry)

DE Rat prolactin-releasing peptide, PRP.

XX Prolactin-releasing peptide; PRP; GPR10; G protein-coupled receptor;
 KW feeding behaviour; food intake; modulation; antagonist; anorectic;
 KW obesity; agonist; cachexia.

OS Rattus sp.

PN WO200017641-A1.

PD 30-MAR-2000.

PF 22-SEP-1999; 99WO-US21243.

PR 22-SEP-1998; 98US-0101380.

PA 14-OCT-1998; 98US-0172353.

PI (MILL-) MILLENNIUM PHARM INC.

PI Stricker-Kongrad A, Gu W;

DR WPI; 2000-303231/26.

PF Identifying modulators of body weight by a combination of a cell-free

PT or cell-based assay to identify modulators of GPR10, followed by an in

PT vivo assay for the compounds effect on e.g. feeding behavior

PS Example 2; Page 61; 82pp; English.

CC The invention relates to a method for identifying compounds useful for
 CC modulating body weight. The method comprises cell-free and/or cell-based
 CC assays that identify compounds which bind to and/or activate or inhibit

CC the activity of GPR10, a G protein-coupled receptor. These assays are
CC then followed by an in vivo assay of the effect of the compound on
CC feeding behaviour, body weight or metabolic rate in a mammal. Prolactin-
CC releasing peptide (PrRP; AAY87504) is a ligand of GPR10. Binding of PrRP
CC to GPR10 stimulates a signal transduction cascade, which results in an
CC increase in food intake. Compounds identified using the method of the
CC invention are useful for the modulation of body weight. Antagonists of
CC GPR10 can be used to treat obesity, while GPR10 agonists can be used to
CC treat cachexia. The present sequence represents rat PrRP.

XX
SQ Sequence 31 AA;

Query Match 91.5%; Score 43; DB 21; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.031;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
Db 1 SRAHQSMETR 11

RESULT 8
AAY49292
ID AAY49292 standard; peptide; 31 AA.

XX
AC AAY49292;

XX
DT 22-FEB-2000 (first entry)

XX
DE 19P2 ligand peptide fragment.

XX
KW Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
XX pituitary; regulatory mechanism; central nervous system; pancreatic.

XX
OS Rattus sp.

XX
FH Key Location/Qualifiers

XX
FT Modified-site 31 /note- "C-terminal amide"

XX
PN WO960112-A1.

XX
PD 25-NOV-1999.

XX
PF 20-MAY-1999; 99WO-JP02650.

XX
PR 21-MAY-1998; 98JP-0140293.

XX
PA (TAKE) TAKEDA CHEM IND LTD.

XX
PI Matsumoto H, Kitada C, Hinuma S;

XX
DR WPI; 2000-039381/03.

XX
PT New monoclonal antibodies, useful in diagnosis, as drugs and in
XX studying diseases related to ligand abnormality

XX
PS Disclosure; Page 26; 73pp; Japanese.

XX
CC The invention provides a monoclonal antibody which has a specific
CC reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC derivative. The antibodies can be used in diagnosis or to treat or
CC prevent diseases associated with abnormality in the pituitary function
CC regulatory mechanism (e.g. promotion of prolactin secretion); central
CC nervous regulatory mechanism, and pancreatic function regulatory
CC mechanism. The antibody-based immunoassay can also be applied in
CC clarifying the physiological functions of the ligand and its derivative.
CC Sequences AAY49290-302 represent peptide fragments of the 19P2 ligand.

XX
SQ Sequence 31 AA;

Query Match 91.5%; Score 43; DB 21; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.031;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
Db 1 SRAHQSMETR 11

RESULT 9
AAG62524
ID AAG62524 standard; peptide; 31 AA.

XX
AC AAG62524;

XX
DT 24-AUG-2001 (first entry)

XX
DE Rat CRH releasing protein related peptide SEQ ID NO: 18.

XX
KW Rat; corticotrophin releasing hormone; CRH; G protein receptor ligand;
KW analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;
KW Addison's disease; adrenal gland hyperfunction; obesity.

XX
OS Rattus sp.

XX
PN WO200135984-A1.

XX
PD 25-MAY-2001.

XX
PF 17-NOV-2000; 2000WO-JP08119.

XX
PR 18-NOV-1999; 99JP-0327900.

XX
PR 26-SEP-2000; 2000JP-0297073.

XX
PA (TAKE) TAKEDA CHEM IND LTD.

XX
PI Kitada C, Matsumoto H, Hinuma S;

XX
DR WPI; 2001-355552/37.

XX
PT Use of G protein receptor ligand or peptide for controlling
XX corticotrophin releasing hormone secretion

XX
PS Claim 3; Page 69; 90pp; Japanese.

XX
CC The present sequence describes a method of controlling the secretion of
CC corticotrophin releasing hormone (CRH), involving the use of a G protein
CC receptor ligand. This can be used to control the secretion of CRH and is
CC useful as an analgesic or for treating, preventing or ameliorating
CC diseases associated with CRH secretion such as hyperaldosteronism,
CC hypercortisolemia, secondary or chronic hypoadrenocorticism, Addison's
CC disease (including boredom, nausea, pigmentation, hypogonadism, hair
CC loss, and hypotension), adrenal gland hypofunction and obesity. The
CC present sequence is a peptide used in the exemplification of the
CC invention.

XX
SQ Sequence 31 AA;

Query Match 91.5%; Score 43; DB 22; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.031;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
Db 1 SRAHQSMETR 11

RESULT 10
AAB90993
ID AAB90993 standard; peptide; 31 AA.

XX
AC AAB90993;

XX
DT 22-JUN-2001 (first entry)

Prolactin releasing peptide SEQ ID NO:167.

KW protection; endogenous therapeutic peptide; peptidase; conjugation;
KW blood component; modification; succinimidyl; maleimido group; amino
KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.

Homo sapiens
Synthetic.

WO200069900-A2

23-NOV-2000.

17-MAY-2000; 2000WO-US13576.

17-MAY-1999; 99US-0134406.
10-SEP-1999; 99US-0153406

15-OCT-1999; 99US-0159783.

(CONF-) CONJUCHEM INC.

Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibault K,

WPI; 2001-112059/12.

Modifying and attaching therapeutic peptides to albumin prevents peptidase degradation, useful for increasing length of *in vivo* activity

xx PS Disclosure; Page 244; 733pp; English.

CC The present invention describes a modified therapeutic peptide (I)

CC comprising a therapeutically active amino acid region (III) and a

CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to

CC a less therapeutically active amino acid region (IV), which covalently

CC bonds with amino/hydroxyl/thiol groups on blood components to form a

CC peptidease stabilised therapeutic peptide composed of 3-50 amino acids.

CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth

CC factors and neurotransmitters, to protect them from peptidease activity

CC in vivo for the treatment of various disorders. Endogenous therapeutic

CC peptides are not suitable as drug candidates as they require frequent

CC administration due to rapid degradation by peptidases in the body.

CC Modifying and attaching therapeutic peptides to albumin prevents or

CC reduces the action of peptideases to increase length of activity (half

CC life) and specificity as bonding to large molecules decreases

CC intracellular uptake and interference with physiological processes.

CC AAB90893 to AAB92441 represent peptides which can be used in the

CC exemplification of the present invention.

Sequence 31 AA;

Query Match	91.5%;	Score 43;	DB 22;	Length 31;
Best Local Similarity	72.7%;	Pred. No. 0.031;		
Matches	8;	Conservative	0;	Mismatches 3;
			Indels 0;	Gaps 0

```
QY 1 SRXHXHSMEXR 11
    ||| ||| |
Db 1 SRAHQHSMETR 11
```

RESULT 11
AAW31385

•AC AAW31385

DT 06-APR-1998 (first entry)

DE Rat type 6 protein-coupled receptor 119 and fragment 2.

KM G protein-coupled receptor; ligand binding; pharmaceutical;
KM modulator; pituitary; central nervous system; pancreas; prophylactic
KM therapeutic agent.

Rat sp

PN WO9724436-A2

PD 10-JUL-1997
YY

PF 26-DEC-1996; 96WO-JP03821.

PR	18-SEP-1996;	96JJP-0246573
PR	28-DEC-1995;	95JJP-0343371

PR 15-MAR-1996; 96JP-0059419
PR 12-ATG-1996; 96JP-0211805

XX
XX
DA (TAKE) TAKEDA CHEM IND LTD

XX Eniti B. Eukusumi S. Hapat

XX Pl Kawamata Y, Kitada C;
XX

DR WPI; 199/-363612/33.
DR N-PSDB: AAV02422.

Ligand peptide for PT XX

PT Ligand-peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
PS Claim 2, Page 179; 258pp; English.
XX

This sentence represents a peptide fragment from a novel rat type I ligand polypeptide corresponding to amino acid residues 22 to 53 of the sequence represented in AM41383 and is used in an assay to monitor ligand binding to the G protein-coupled receptor protein. Pharmaceutical compositions containing this ligand may be used as a pituitary function modulator, a central nervous system modulator or a paracrine function modulator. This ligand could have specific applications as a prophylactic or therapeutic agent for dementia, depression, hyperkinetic syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia, hypercholesterolaemia, hyperglycaemia, hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease, Turner's syndrome, neuritis, asthma, rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis, acute myocardial infarction, infertility, spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis, osteoporosis and/or oligosaccharia. Assays can also be developed to screen compounds which are capable of altering the binding activity of the ligand affecting activation of the G protein-coupled receptor protein.

SQ	Sequence	32 AA
----	----------	-------

Query Match	91.5%;	Score 43;	DB 18;	Length 32;
Best Local Similarity	72.7%;	Pred. No. 0.032;		
Matches	8;	Conservative	0;	Mismatches 3;
			Indels	0;
			Gaps	0;

```

OY      1 SRXHXHSMEXR 11
        ||| ||||| |
Db      1 SRAHQHSMETR 11

```

RESULT	12
AAB10356	
ID	AAB10356 standard; peptide; 32 AA

AAC10356;

DT 24-NOV-2000 (first entry)

DE Rat oxytocin secretion promoting peptide SEQ ID NO: 19

KM Rat; oxytocin secretion promoter; G protein-coupled receptor protein;
KM treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KM caesarean section; artificial fertilization; galactostasis; goat; pig;
KM veterinary medicine; milk production.

OS	Rattus sp.
XX	
PN	WO200038704-A1.
XX	
PD	06-JUL-2000.
PF	22-DEC-1999; 99WO-JP07199.
XX	
PR	25-DEC-1998; 98JP-0369585.
XX	
PA	(TAKE) TAKEDA CHEM IND LTD.
XX	
PI	Matsumoto H, Kitada C, Hinuma S;
XX	
DR	WPI: 2000-452298/39.
XX	
PT	Physiologically-active polypeptide recognized as ligand by G
PT	protein-coupled receptor protein, for promoting secretion of oxytocin,
PT	as drugs for diseases relating to oxytocin secretion and in veterinary
PT	medicine
XX	
PS	Disclosure; Page 57; 72pp; Japanese.
XX	
CC	This invention describes a novel oxytocin secretion-regulating agent
CC	which contains a ligand peptide or its salt for the G protein-coupled
CC	receptor protein. It is useful in the form of drugs for ameliorating,
CC	preventing and treating diseases relating to oxytocin secretion e.g.,
CC	wetle pains and atonic bleeding, before and after expulsion of placenta,
CC	uterine recovery failure, caesarean section, stoppage of artificial
CC	fertilization or galactostasis and is also applicable in veterinary
CC	medicine for promoting milk production in cow, goat and pig. This
CC	sequence represents a rat peptide which acts as an oxytocin secretion
CC	promoter.
XX	
SQ	Sequence 32 AA:
OY	1 SRXHXSMEXR 11 1 SHAOHSMETR 11
DB	
RESULT 13	
AAG62525	
ID	AAG62525 standard; peptide: 32 AA.
XX	
AC	AAG62525;
XX	
DJ	24-AUG-2001 (first entry)
DE	
XX	Rat CRH releasing protein related peptide SEQ ID NO: 19.
XX	
KW	Rat; corticotrophin releasing hormone; CRH; G protein receptor ligand;
KW	analgesic; hyperaldosteronism; hypercortisolaemia; hypoadrenocorticism;
KW	Addison's disease; adrenal gland hyperfunction; obesity.
XX	
OS	Rattus sp.
XX	
PN	WO200135984-A1.
XX	
PD	25-MAY-2001.
XX	
PF	17-NOV-2000; 2000WO-JP08119.
XX	
PR	18-NOV-1999; 99JP-0372900.
PR	26-SEP-2000; 2000JP-0297073.
XX	
PA	(TAKE) TAKEDA CHEM IND LTD.
XX	
PI	Kitada C, Matsumoto H, Hinuma S;

XX	WPI, 2001-355552/37.
DR	
XX	
PT	Use of G protein receptor ligand or peptide for controlling
PT	corticotropin releasing hormone secretion -
XX	
PS	Disclosure; Page 69; 90pp; Japanese.
XX	
CC	The present sequence describes a method of controlling the secretion of
CC	corticotrophin releasing hormone (CRH), involving the use of a G protein
CC	receptor ligand. This can be used to control the secretion of CRH and is
CC	useful as an analgesic or for treating, preventing or ameliorating
CC	diseases associated with CRH secretion such as hyperaldosteronism,
CC	hypocortisolemia, secondary or chronic hypoadrenocorticism, Addison's
CC	disease (including, boredom, nausea, pigmentation, hypogonadism, hair
CC	loss, and hypotension), adrenal gland hypofunction and obesity. The
CC	present sequence is a peptide used in the exemplification of the
CC	invention.
XX	
SO	Sequence 32 AA;
Query Match	91.5%; Score 43; DB 22; Length 32;
Best Local Similarity	72.7%; Pred. No. 0.032;
Matches, 8; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
OY	1 SRXHXSMEXR 11
Db	1 SRAHQSMETR 11
RESULT 14	
AAW31386	
ID	AAW31386 standard; Peptide; 33 AA.
XX	
AC	AAW31386;
XX	
DT	06-APR-1998 (first entry)
XX	
DE	Rat type G protein-coupled receptor ligand fragment 3.
XX	
KW	G protein-coupled receptor; ligand binding; pharmaceutical;
KW	modulator; pituitary; central nervous system; pancreas; prophylactic;
KW	therapeutic agent.
XX	
OS	Rat sp.
XX	
PN	WO9724436-A2.
XX	
PD	10-JUL-1997.
XX	
PF	26-DEC-1996; 96WO-JP03821.
XX	
PR	18-SEP-1996; 96JP-0246573.
PR	28-DEC-1995; 95JP-0343371.
PR	15-MAR-1996; 96JP-0059419.
PR	12-AUG-1996; 96JP-0211805.
XX	
PA	(TAKE) TAKEDA CHEM IND LTD.
XX	
PI	Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
PI	Kawamata Y, Kitada C;
XX	
DR	WPI, 1997-363672/33.
XX	
DR	N-PSDB; AAV02423.
XX	
PS	Claim 2; Page 179-180; 258pp; English.
CC	This sequence represents a peptide fragment from a novel rat type
CC	ligand polypeptide corresponding to amino acid residues 22 to 54 of the
CC	sequence represented in AAW31383 and is used in an assay to monitor

CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atrophic dermatitis, osteoporosis and/or
 CC oligogalactia. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 CC
 SQ Sequence 33 AA;

Query Match 91.5%; Score 43; DB 18; Length 33;
 Best Local Similarity 72.7%; Pred. No. 0.033;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
 || || || || ||
 Db 1 SRAHQSMETR 11

RESULT 15
 AAB10357

ID AAB10357 standard; peptide; 33 AA.

AC AAB10357;

DF 24-NOV-2000 (first entry)

DE Rat oxytocin secretion promoting peptide SEQ ID NO: 20.

XX Rat; oxytocin secretion promoter; G protein-coupled receptor protein;

KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;

KW caesarean section; artificial fertilization; galactostasis; goat; pig;

XX veterinary medicine; milk production.

OS Rattus sp.

PN W0200038704-A1.

PD 06-JUL-2000.

PF 22-DEC-1999; 99WO-JP07199.

PR 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-452298/39.

PT Physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine

PS Disclosure; Page 58; 72pp; Japanese.

XX This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary

CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a rat peptide which acts as an oxytocin secretion
 CC promoter.

XX SQ Sequence 33 AA;

Query Match 91.5%; Score 43; DB 21; Length 33;
 Best Local Similarity 72.7%; Pred. No. 0.033;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
 || || || || ||
 Db 1 SRAHQSMETR 11

Search completed: May 1, 2003, 14:30:50
 Job time : 15.2969 secs

File Copy
Thu May 1 15:54:02 2003
Cited in paper #16
de-Run

seq74-plus-73.rag

SEQ ID NO: 74 fused to SEQ ID NO: 73
Database: A-GeneSeq-101002
AC NO: AAW31391
Page 1

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:36:27 ; Search time 35 Seconds
(without alignments)
125.636 Million cell updates/sec

Title: SEQ74-PLUS-73
Perfect score: 156
Sequence: 1 SRKXHSKEXRTPDINPAMVYKXRGIRPVGRFX 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA2001.DAT.*
 - 23: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	94.2	31	AAW31391	Human type G prote
2	147	94.2	31	AAW31384	Rat type G protein
3	147	94.2	31	AAW97233	Rat type G protein
4	147	94.2	31	AAW97233	Rat type G protein
5	147	94.2	31	AAW97233	Rat type G protein
6	147	94.2	31	AAW97233	Rat type G protein
7	147	94.2	31	AAW97233	Rat type G protein
8	147	94.2	31	AAW97233	Rat type G protein
9	147	94.2	31	AAW97233	Rat type G protein
10	147	94.2	31	AAW97233	Rat type G protein

11	147	94.2	31	AAW97504	Rat prolactin-rele
12	147	94.2	31	AAW49291	19P2 ligand peptid
13	147	94.2	31	AAW49292	19P2 ligand peptid
14	147	94.2	31	AAW62524	Rat CRH releasing
15	147	94.2	31	AAW62531	Human CRH releasing
16	147	94.2	31	AAW90991	Human CRH releasing
17	147	94.2	31	AAW90993	Proactin releasin
18	147	94.2	31	AAW90995	Proactin releasin
19	147	94.2	31	AAW31392	Human type G prote
20	147	94.2	31	AAW31385	Rat type G protein
21	147	94.2	31	AAW10356	Rat oxytocin secre
22	147	94.2	31	AAW10357	Human oxytocin sec
23	147	94.2	31	AAW62525	Rat CRH releasing
24	147	94.2	31	AAW62532	Human CRH releasin
25	147	94.2	31	AAW31393	Human type G prote
26	147	94.2	31	AAW31386	Rat type G protein
27	147	94.2	31	AAW10357	Rat oxytocin secre
28	147	94.2	31	AAW10357	Human oxytocin sec
29	147	94.2	31	AAW10357	Rat CRH releasing
30	147	94.2	31	AAW62532	Human CRH releasin
31	147	94.2	31	AAW95172	Human CRH releasin
32	147	94.2	31	AAW31383	Murine pituitary-d
33	147	94.2	31	AAW97225	Rat type G protein
34	147	94.2	31	AAW97225	Rat type G protein
35	147	94.2	31	AAW10354	Rat type G protein
36	147	94.2	31	AAW62523	Rat oxytocin secre
37	147	94.2	31	AAW31390	Human type G prote
38	147	94.2	31	AAW97226	Human type G prote
39	147	94.2	31	AAW10361	Human oxytocin sec
40	147	94.2	31	AAW62530	Human CRH releasin
41	146	93.6	31	AAW31371	Bovine G protein-c
42	146	93.6	31	AAW97218	Bovine pituitary-d
43	146	93.6	31	AAW97218	Bovine pituitary-d
44	146	93.6	31	AAW95188	Bovine pituitary-d
45	146	93.6	31	AAW95188	Bovine pituitary-d

ALIGNMENTS

RESULT 1	AAW31391	AAW31391 standard; Peptide: 31 AA.
ID	AAW31391	
AC	AAW31391	
DT	06-APR-1998	(first entry)
XX		Human type G protein-coupled receptor ligand fragment 1.
DE		G protein-coupled receptor; ligand binding; pharmaceutical;
KW		modulator; pituitary; central nervous system; pancreas; prophylactic;
KW		therapeutic agent.
XX		
OS	Homo sapiens.	
PN	W09724436-A2.	
PD	10-JUL-1997.	
XX		
FE	26-DEC-1996;	96MO-JP03821.
XX		
PR	18-SEP-1996;	96GP-0246573.
PR	28-DEC-1995;	95JP-0343371.
PR	15-MAR-1996;	96GP-0059419.
XX	12-AUG-1996;	96GP-0211805.
XX		
PA	(TAKE) TAKEDA CHEM IND LTD.	
XX		
PI	Fujita R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;	
PI	Kawamata Y, Kitada C;	
XX		
DR	WPI; 1997-363672/33.	

DR N-PSDB; AAV02428.

XX Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX

PS Claim 2; Page 184; 258pp; English.

XX This sequence represents a peptide fragment from a novel human type
CC ligand polypeptide corresponding to amino acid residues 23 to 53 of the
CC sequence represented in AAW31384 and is used in an assay to monitor
CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a
CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
CC trauma, growth hormone secretory disease, hyper- and polypnagia,
CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
CC Turner's syndrome, neurosis, asthma, amyotrophic lateral sclerosis,
CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
CC acute myocardial infarction, infertility, spinocerebellar degeneration,
CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
CC oligogalactia. Assays can also be developed to screen compounds which are
CC capable of altering the binding activity of the ligand affecting
CC activation of the G protein-coupled receptor protein.

XX Sequence 31 AA;

Query Match 94.2%; Score 147; DB 18; Length 31;

Best Local Similarity 83.9%; Pred. No. 1.7e-11;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMVYXXRGIRPVGRF 31

DB 1 SRTHRSMETRTPDINPAMVYASRGIRPVGRF 31

RESULT 2

AAW31384
ID AAW31384 standard; peptide; 31 AA.

XX AAW31384;

DT 06-APR-1998 (first entry)

XX Rat type G protein-coupled receptor ligand fragment 1.

XX G protein-coupled receptor; ligand binding; pharmaceutical;
KW modulator; pituitary; central nervous system; pancreas; prophylactic;
KW therapeutic agent.

XX Rat sp.

XX WO9724436-A2.

XX 10-JUL-1997.

XX 26-DEC-1996; 96WO-JP03821.

XX 18-SEP-1996; 96JP-0246573.

XX 28-DEC-1995; 95JP-0343371.

XX 15-MAR-1996; 96JP-0059419.

XX 12-AUG-1996; 96JP-0211805.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;

XX Kawamata Y, Kitada C;

XX WPI: 1997-363672/33.

XX N-PSDB; AAV02421.

PT Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX

PS Claim 2; Page 179; 258pp; English.

XX This sequence represents a peptide fragment from a novel rat type
CC ligand polypeptide corresponding to amino acid residues 22 to 52 of the
CC sequence represented in AAW97233 and is used in an assay to monitor
CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a
CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
CC trauma, growth hormone secretory disease, hyper- and polypnagia,
CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
CC hyperprolactinaemia, diabetes, cancer, pancreatitis, renal disease,
CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
CC acute myocardial infarction, infertility, spinocerebellar degeneration,
CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
CC oligogalactia. Assays can also be developed to screen compounds which are
CC capable of altering the binding activity of the ligand affecting
CC activation of the G protein-coupled receptor protein.

XX Sequence 31 AA;

Query Match 94.2%; Score 147; DB 18; Length 31;

Best Local Similarity 83.9%; Pred. No. 1.7e-11;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMVYXXRGIRPVGRF 31

DB 1 SRAHOSMETRTPDINPAMVYTGIRPVGRF 31

RESULT 3

AAW97233
ID AAW97233 standard; peptide; 31 AA.

XX AAW97233;

DT 06-MAY-1999 (first entry)

XX Rat type ligand polypeptide fragment.

XX Rat type ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypovarianism; gonocyst cacosgenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmenopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; chorioncarcinoma; hydralid mole;
KW abortion; unfertilized fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.

XX Rattus sp.

XX WO9858962-A1.

XX 30-DEC-1998.

XX 22-JUN-1998; 98WO-JP02765.

XX 23-JUN-1997; 97JP-0165437.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Fujii R, Hinuma S, Kawamata Y, Matsumoto H;

XX WPI: 1999-105614/09.

PT Use of G protein-coupled receptor ligands - for modulating prolactin
 secretion or placental function, e.g. for treating menopausal
 syndrome, tumours, autoimmune disease or abnormal pregnancy

PS Claim 3; Page 153; 241pp; English.

CC The present sequence represents a rat type ligand fragment. It
 CC is used in the course of the invention. The specification describes
 CC an agent for modulating prolactin secretion which comprises a
 CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
 CC protein. The agents for promoting prolactin secretion can be used for
 CC treating or preventing hypovarianism, gonocyst cocogenesis, menopausal
 CC syndrome, euthyroid or hypometabolism. They can be used for promoting
 CC lactation in a domestic mammal and as an aphrodisiac. The agents for
 CC inhibiting prolactin secretion can be used for treating or preventing
 CC pituitary adenomatosis, brain tumour, amenorrhoea, galactorrhea,
 CC prolactinoma, infertility, impotence, emmenorrhoea, autoimmune disease,
 CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
 CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
 CC The inhibitory agents can also be used as contraceptives. The agents for
 CC modulating placental function can be used for treating or preventing
 CC chorioncarcinoma, hydatid mole, abortion, uterine fetus,
 CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.

SQ Sequence 31 AA:

Query Match 94.2%; Score 147; DB 20; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1.7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
 DB 1 SRAHQSMETRTPDINPAMYTGKIRPVGRF 31

RESULT 4

AAW97235
 ID AAW97235 standard; peptide: 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

KW Rat type ligand; modulation; prolactin secretion;
 KW G protein-coupled receptor; GPCR; hypovarianism; gonocyst cocogenesis;
 KW menopausal syndrome; euthyroid; hypometabolism; lactation;
 KW pituitary adenomatosis; brain tumour; amenorrhoea; autoimmune disease;
 KW prolactinoma; infertility; impotence; emmenorrhoea; galactorrhea;
 KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
 KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
 KW contraceptive; placental function; chorioncarcinoma; hydatid mole;
 KW abortion; uterine fetus; abnormal saccharometabolism;
 KW abnormal lipidmetabolism; oxytocia.

OS Homo sapiens.

PN WO9858962-A1.

PD 30-DEC-1998

PF 22-JUN-1998; 98WO-JP02765.

PR 23-JUN-1997; 97JP-0165437.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;

WPI: 1999-105614/09.

Use of G protein-coupled receptor ligands - for modulating prolactin

PT secretion or placental function, e.g. for treating menopausal
 syndrome, tumours, autoimmune disease or abnormal pregnancy

PS Claim 3; Page 159; 241pp; English.

CC The present sequence represents a human type ligand fragment. It
 CC is used in the course of the invention. The specification describes
 CC an agent for modulating prolactin secretion which comprises a
 CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
 CC protein. The agents for promoting prolactin secretion can be used for
 CC treating or preventing hypovarianism, gonocyst cocogenesis, menopausal
 CC syndrome, euthyroid or hypometabolism. They can be used for promoting
 CC lactation in a domestic mammal and as an aphrodisiac. The agents for
 CC inhibiting prolactin secretion can be used for treating or preventing
 CC pituitary adenomatosis, brain tumour, amenorrhoea, autoimmune disease,
 CC prolactinoma, infertility, impotence, emmenorrhoea, galactorrhea,
 CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
 CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
 CC The inhibitory agents can also be used as contraceptives. The agents for
 CC modulating placental function can be used for treating or preventing
 CC chorioncarcinoma, hydatid mole, abortion, uterine fetus,
 CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.

SQ Sequence 31 AA:

Query Match 94.2%; Score 147; DB 20; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1.7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
 DB 1 SRTHRSMETRTPDINPAMYASRGIRPVGRF 31

RESULT 5

AAW87614
 ID AAW87614 standard; Peptide: 31 AA.

AC AAW87614;

DT 29-MAR-1999 (first entry)

DE Rat 19P2 ligand.

KW 19P2 ligand; G protein coupled receptor; pituitary;
 KW prolactin releasing peptide; rat; dementia; breast cancer;
 KW therapy.

OS Rattus sp.

PN EP887417-A2.

PD 30-DEC-1998.

PF 25-JUN-1998; 98EP-0111725.

PR 27-JUN-1997; 97JP-0172118.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;

WPI: 1999-047884/05.

Producing a 19P2 pituitary G protein receptor ligand - by cleavage
 of a fusion protein, useful for preventing and treating dementia,
 breast cancer, renal failure and autoimmune disease

Claim 5; Page 34; 56pp; English.

This is the amino acid sequence of the rat pituitary G
 protein-coupled receptor ligand 19P2. A method suitable for
 commercial high-level production of 19P2L comprises expressing

PT physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine
 PS Claim 3; Page 57; 72pp; Japanese.
 XX
 CC This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.,
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a rat peptide which acts as an oxytocin secretion
 CC promoter.
 CC
 SQ Sequence 31 AA;
 Query Match 94.2%; Score 147; DB 21; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1,7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 1 SRXHSMEKRTPDINPAWYXXRGIRPVGRF 31
 DB 1 SRAHQHSMERTPTDINPAWYTGRIKRVGRF 31
 RESULT 10
 AAB10362
 ID AAB10362 standard; peptide; 31 AA.
 AC AAB10362;
 XX
 DT 24-NOV-2000 (first entry)
 DE Human oxytocin secretion promoting peptide SEQ ID NO: 32.
 XX
 KW Human; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.
 XX
 OS Homo sapiens.
 XX
 PN WO200038704-A1.
 XX
 PD 06-JUL-2000.
 XX
 PF 22-DEC-1999; 99WO-JP07199.
 XX
 PR 25-DEC-1998; 98JP-0369585.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Matsumoto H, Kitada C, Hinuma S;
 XX
 DR WPI; 2000-452298/39.
 XX
 PT Physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine
 PS Disclosure; Page 62; 72pp; Japanese.
 XX
 CC This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.,
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial

CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a human peptide which acts as an oxytocin secretion
 CC promoter.
 CC
 SQ Sequence 31 AA;
 Query Match 94.2%; Score 147; DB 21; Length 31;
 Best Local Similarity 83.9%; Pred. No. 1,7e-11;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 1 SRXHSMEKRTPDINPAWYXXRGIRPVGRF 31
 DB 1 SRTHRHSMERTPTDINPAWYASRGIRPVGRF 31
 RESULT 11
 AAY87504
 ID AAY87504 standard; protein; 31 AA.
 AC AAY87504;
 XX
 DT 18-JUL-2000 (first entry)
 DE Rat prolactin-releasing peptide, PrRP.
 XX
 DE Prolactin-releasing peptide; PrRP; GPR10; G protein-coupled receptor;
 KW feeding behaviour; food intake; modulation; antagonist; anorectic;
 KW obesity; agonist; cachexia.
 XX
 OS Rattus sp.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 31
 FT /note="C-terminal amide"
 XX
 PN WO200017641-A1.
 XX
 PD 30-MAR-2000.
 XX
 PF 22-SEP-1999; 99WO-US21243.
 XX
 PR 22-SEP-1998; 98US-0101380.
 PR 14-OCT-1996; 96US-0172353.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Stricker-Kongrad A, Gu W;
 XX
 DR WPI; 2000-303231/26.
 XX
 PT Identifying modulators of body weight by a combination of a cell-free
 PT or cell-based assay to identify modulators of GPR10, followed by an in
 PT vivo assay for the compounds effect on e.g. feeding behavior
 XX
 PS Example 2; Page 61; 82pp; English.
 XX
 CC The invention relates to a method for identifying compounds useful for
 CC modulating body weight. The method comprises cell-free and/or cell-based
 CC assays that identify compounds which bind to and/or activate or inhibit
 CC the activity of GPR10, a G protein-coupled receptor. These assays are
 CC then followed by an in vivo assay of the effect of the compound on
 CC feeding behaviour, body weight or metabolic rate in a mammal. Prolactin-
 CC releasing peptide (PrRP; AAY87504) is a ligand of GPR10. Binding of PrRP
 CC to GPR10 stimulates a signal transduction cascade, which results in an
 CC increase in food intake. Compounds identified using the method of the
 CC invention are useful for the modulation of body weight. Antagonists of
 CC GPR10 can be used to treat obesity, while GPR10 agonists can be used to
 CC treat cachexia. The present sequence represents rat PrRP.
 CC
 SQ Sequence 31 AA;
 Query Match 94.2%; Score 147; DB 21; Length 31;

Best Local Similarity 83.9%; Pred. No. 1,7e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRAHQSMETRTPDINPAMYGRGIRPVGRF 31

RESULT 12

AA49291
ID AAY49291 standard; peptide; 31 AA.

AC AAY49291;

DT 22-FEB-2000 (first entry).

DE 19P2 ligand peptide fragment.

KW Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW pituitary; regulatory mechanism; central nervous system; pancreatic.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 31 /note="C-terminal amide"

XX WO960112-A1.

PD 25-NOV-1999.

XX 20-MAY-1999; 99WO-JP02650.

PR 21-MAY-1998; 98JP-0140293.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Matsumoto H, Kitada C, Hinuma S;

XX WPI; 2000-039381/03.

DR WPI; 2000-039381/03.

XX New monoclonal antibodies, useful in diagnosis, as drugs and in

PT studying diseases related to ligand abnormality

XX Disclosure; Page 26; 73pp; Japanese.

PS The invention provides a monoclonal antibody which has a specific

CC reaction with the part peptide of the C-terminal of 19P2 ligand or its

CC derivative. The antibodies can be used in diagnosis or to treat or

CC prevent diseases associated with abnormality in the pituitary function

CC regulatory mechanism (e.g. promotion of prolactin secretion), central

CC nervous regulatory mechanism, and pancreatic function regulatory

CC mechanism. The antibody-based immunosay can also be applied in

CC clarifying the physiological functions of the ligand and its derivative.

CC Sequences AAY49290-302 represent peptide fragments of the 19P2 ligand.

XX Sequence 31 AA;

SO Query Match 94.2%; Score 147; DB 21; Length 31;

Best Local Similarity 83.9%; Pred. No. 1.7e-11;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31

DB 1 SRAHQSMETRTPDINPAMYGRGIRPVGRF 31

RESULT 13
AAY49292
ID AAY49292 standard; peptide; 31 AA.
AC AAY49292;
XX

DT 22-FEB-2000 (first entry)

XX 19P2 ligand peptide fragment.

DE Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;

KW pituitary; regulatory mechanism; central nervous system; pancreatic.

XX Rattus sp.

XX Key Location/Qualifiers

FT Modified-site 31 /note="C-terminal amide"

XX WO960112-A1.

PD 25-NOV-1999.

XX 20-MAY-1999; 99WO-JP02650.

PR 21-MAY-1998; 98JP-0140293.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Matsumoto H, Kitada C, Hinuma S;

XX WPI; 2000-039381/03.

DR WPI; 2000-039381/03.

XX New monoclonal antibodies, useful in diagnosis, as drugs and in

PT studying diseases related to ligand abnormality

XX Disclosure; Page 26; 73pp; Japanese.

PS The invention provides a monoclonal antibody which has a specific

CC reaction with the part peptide of the C-terminal of 19P2 ligand or its

CC derivative. The antibodies can be used in diagnosis or to treat or

CC prevent diseases associated with abnormality in the pituitary function

CC regulatory mechanism (e.g. promotion of prolactin secretion), central

CC nervous regulatory mechanism, and pancreatic function regulatory

CC mechanism. The antibody-based immunosay can also be applied in

CC clarifying the physiological functions of the ligand and its derivative.

CC Sequences AAY49290-302 represent peptide fragments of the 19P2 ligand.

XX Sequence 31 AA;

SO Query Match 94.2%; Score 147; DB 21; Length 31;

Best Local Similarity 83.9%; Pred. No. 1.7e-11;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31

DB 1 SRAHQSMETRTPDINPAMYGRGIRPVGRF 31

RESULT 14
AAG62524
ID AAG62524 standard; peptide; 31 AA.

AC AAG62524;

DT 24-AUG-2001 (first entry)

DE Rat CRH releasing protein related peptide SEQ ID NO: 18.

XX Rat; corticotrophin releasing hormone; CRH; G protein receptor ligand;

KW analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;

XX Addison's disease; adrenal gland hyperfunction; obesity.

OS Rattus sp.
XX WO200135984-A1.
XX 25-MAY-2001.
XX

Thu May 1 15:54:14 2003

De-Run

us-09-446-543a-61.rag

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 43.1094 Seconds
(without alignments)
95.821 Million cell updates/sec

Title: US-09-446-543a-61
Perfect score: 171
Sequence: 1 SKRHSMETKTPDINPAMVSRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
V1	171	100.0	31	AAW31391	Human type G prote
2	171	100.0	31	AAW97235	Human type G prote
3	171	100.0	31	AAW87615	Human type G prote
4	171	100.0	31	AAW10362	Human type G prote
5	171	100.0	31	AAW49291	Human type G prote
6	171	100.0	31	AAW62531	Human type G prote
7	171	100.0	31	AAW90991	Human type G prote
8	171	100.0	31	AAW31392	Human type G prote
9	171	100.0	32	AAW10363	Human type G prote
10	171	100.0	32	AAW62532	Human type G prote

See ID NO: 61
Ae NO: AAW31391
Database: A-Geneseg-101002
File Copy
Page 1
Cited in paper

11	171	100.0	33	AAW31393	Human type G prote
12	171	100.0	33	AAW10364	Human type G prote
13	171	100.0	33	AAW62533	Human type G prote
14	171	100.0	37	AAW31390	Human type G prote
15	171	100.0	87	AAW97226	Human type G prote
16	171	100.0	87	AAW10361	Human type G prote
17	171	100.0	87	AAW62530	Human type G prote
18	165	96.5	30	AAW49292	Human type G prote
19	162	94.7	31	AAW90995	Human type G prote
20	158	92.4	31	AAW31371	Human type G prote
21	158	92.4	31	AAW97218	Human type G prote
22	158	92.4	31	AAW87613	Human type G prote
23	158	92.4	31	AAW95188	Human type G prote
24	158	92.4	31	AAW10347	Human type G prote
25	158	92.4	31	AAW49290	Human type G prote
26	158	92.4	31	AAW49298	Human type G prote
27	158	92.4	31	AAW62516	Human type G prote
28	158	92.4	32	AAW31372	Human type G prote
29	158	92.4	32	AAW95189	Human type G prote
30	158	92.4	32	AAW10348	Human type G prote
31	158	92.4	32	AAW62517	Human type G prote
32	158	92.4	33	AAW31373	Human type G prote
33	158	92.4	33	AAW95190	Human type G prote
34	158	92.4	33	AAW10349	Human type G prote
35	158	92.4	33	AAW49297	Human type G prote
36	158	92.4	33	AAW62518	Human type G prote
37	158	92.4	33	AAW31382	Human type G prote
38	158	92.4	38	AAW31368	Human type G prote
39	158	92.4	98	AAW97224	Human type G prote
40	158	92.4	98	AAW97217	Human type G prote
41	158	92.4	98	AAW95187	Human type G prote
42	158	92.4	98	AAW10346	Human type G prote
43	158	92.4	98	AAW10353	Human type G prote
44	158	92.4	98	AAW62515	Human type G prote
45	158	92.4	98	AAW62522	Human type G prote

ALIGNMENTS

RESULT 1
ID AAW31391 standard; Peptide: 31 AA.
AC AAW31391;
DE 06-APR-1998 (first entry)
DE Human type G protein-coupled receptor ligand fragment 1.
DE G protein-coupled receptor; ligand binding; pharmaceutical;
DE modulator; pituitary; central nervous system; pancreas; prophylactic;
DE therapeutic agent.
OS Homo sapiens.
XX W09724436-A2
XX 10-JUL-1997.
XX 26-DEC-1996; 96W0-JP03821.
XX 18-SEP-1996; 96JP-0246573.
XX 28-DEC-1995; 95JP-0343371.
XX 15-MAR-1996; 96JP-0059419.
XX 12-AUG-1996; 96JP-0211805.
XX (TAKE) TAKEDA CHEM IND LTD.
XX Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M,
XX Kawamata Y, Kitada C,
XX WPI: 1997-363672/33.

DR N-PSDB; AAV02428.

XX Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
XX Claim 2; Page 184; 258pp; English.

CC This sequence represents a peptide fragment from a novel human type
CC ligand polypeptide corresponding to amino acid residues 23 to 53 of the
CC sequence represented in AAW97235 and is used in an assay to monitor
CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator. A central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a
CC prophylactic or therapeutic agent for dementia, depression, hyperkinetic
CC syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
CC trauma, growth hormone secretory disease, hyperglycemia,
CC hyperlipidaemia, hypercholesterolemia, hyperglycemia,
CC hyperprolactinaemia, diabetes, cancer, pancreaticitis, renal disease,
CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
CC acute myocardial infarction, infertility, spinocerebellar degeneration,
CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
CC oligogalacta. Assays can also be developed to screen compounds which are
CC capable of altering the binding activity of the ligand affecting
CC activation of the G protein-coupled receptor protein.

SQ Sequence 31 AA;

Query Match 100.0%; Score 171; DB 18; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.9e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTPDINPAMWASRGIRPVGRF 31
DB 1 SRTHRSMETRTPDINPAMWASRGIRPVGRF 31

RESULT 2
AAW97235
ID AAW97235 standard; peptide; 31 AA.

AC AAW97235;

DT 06-MAY-1999 (first entry)

DE Human type ligand polypeptide fragment.

XX Rat type ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypocoarctanism; gonocyst cacogenesis;
KW menopausal syndrome; euthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmenorrhoea; galactorrhea;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW interruption mole; abortion; unfertilized fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.

XX Homo sapiens.

XX OS W09858962-A1.

XX PN 30-DEC-1998.

XX PD 22-JUN-1998; 98WO-JP02765.

XX PR 23-JUN-1997; 97JP-0165437.

XX PA (TAKE) TAKEDA CHEM IND LTD.

XX FUJII R, Hinuma S, Kawamata Y, Matsumoto H;
XX

DR WPI; 1999-105614/09.

XX Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
XX Claim 3; Page 159; 241pp; English.

CC The present sequence represents a human type ligand fragment. It
CC is used in the course of the invention. The specification describes
CC an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypocoarctanism, gonocyst cacogenesis, menopausal
CC syndrome, euthyroid or hypometabolism. They can be used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmenorrhoea, galactorrhea,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, abortion, unfertilized fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.

SQ Sequence 31 AA;

Query Match 100.0%; Score 171; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.9e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTPDINPAMWASRGIRPVGRF 31
DB 1 SRTHRSMETRTPDINPAMWASRGIRPVGRF 31

RESULT 3
AAW87615
ID AAW87615 standard; peptide; 31 AA.

AC AAW87615;

DT 29-MAR-1999 (first entry)

DE Human 19P2 ligand.

XX 19P2 ligand; G protein coupled receptor; pituitary;
KW prolactin releasing peptide; human; dementia; breast cancer;
KW therapy.

XX Homo sapiens.

XX EP887417-A2.

XX PD 30-DEC-1998.

XX PF 25-JUN-1998; 98EP-0111725.

XX PR 27-JUN-1997; 97JP-0172118.

XX PA (TAKE) TAKEDA CHEM IND LTD.

XX MORIYA T, Nishimura O, Suenaga M, Tanaka Y;

XX WPI; 1999-047884/05.

XX Producing a 19P2 pituitary G protein receptor ligand - by cleavage
PT of a fusion protein, useful for preventing and treating dementia,
PT breast cancer, renal failure and autoimmune disease
XX
XX Claim 5; Page 35; 56pp; English.

This is the amino acid sequence of the human pituitary G protein-coupled receptor ligand 19P2. A method suitable for commercial high-level production of 19P2L comprising expressing the ligand in host cells as a recombinant fusion protein e.g. with human basic fibroblast growth factor (see AAV83796-97) that has been modified to include an N-terminal cysteine residue. The ligand is released from the fusion by cyanilation followed by ammonolysis. 19P2L has prolactin secretion-stimulating and (at high doses) prolactin secretion-inhibiting properties. It can be used in the treatment and prevention of various diseases including: senile dementia, cerebrovascular dementia, and dementia associated with: neurological disorders (e.g. Alzheimer's disease, Parkinson's disease, Pick's disease, Huntington's disease), infectious diseases (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism, intoxication by drugs, metal and organic compounds), tumorigenic diseases (e.g. brain tumour), traumatic diseases (e.g. chronic subarachnoid haemorrhage, and other types of dementia, depression, hyperactive child syndrome (microencephalopathy) and disturbance of consciousness. It is also useful for prevention and treatment of diseases associated with prolactin hypo and hypersecretion respectively, including: hyperprolactinaemia, pituitary adenoma, breast cancer, infertility, impotence and autoimmune disease (hypersecretion disorders), and seminal vesicle hypoplasia, osteoporosis, menopausal syndrome and renal failure (hyposecretion disorders). The 19P2 polypeptide/amide is also useful as a reagent for study of the prolactin secretory function or as a lactogogue in mammalian farm animals.

Sequence 31 AA;

Query Match 100.0%; Score 171; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.9e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTPDINPAMYASRGIRPVGRF 31
|||||
DB 1 SRTHRSMETRTPDINPAMYASRGIRPVGRF 31

RESULT 4
AAB10362
ID AAB10362 standard; peptide: 31 AA.

AC AAB10362;

DT 24-NOV-2000 (first entry)

DE Human oxytocin secretion promoting peptide SEQ ID NO: 32.

KM Human; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KM caesarean section; artificial fertilization; galactostasis; goat; pig;
KM veterinary medicine; milk production.

OS Homo sapiens.

PN WO200038704-A1.

PD 06-JUL-2000.

PF 22-DEC-1999; 99WO-JP07199.

PR 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-452298/39.

PT Physiologically-active polypeptide recognized as ligand by G
protein-coupled receptor protein, for promoting secretion of oxytocin.

PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine
XX
PS Disclosure; Page 62; 72pp; Japanese.

CC This invention describes a novel oxytocin secretion-regulating agent
CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a human peptide which acts as an oxytocin secretion
CC promoter.

Sequence 31 AA;

Query Match 100.0%; Score 171; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.9e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTPDINPAMYASRGIRPVGRF 31
|||||
DB 1 SRTHRSMETRTPDINPAMYASRGIRPVGRF 31

RESULT 5
AAV49291
ID AAV49291 standard; peptide: 31 AA.

AC AAV49291;

DT 22-FEB-2000 (first entry)

DE 19P2 ligand peptide fragment.

KM Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
KW pituitary; regulatory mechanism; central nervous system; pancreatic.

OS Homo sapiens.

FT Key Location/Qualifiers

FT Modified-site 31 /note="C-terminal amide"

PN WO960112-A1.

PD 25-NOV-1999.

PF 20-MAY-1999; 99WO-JP02650.

PR 21-MAY-1998; 98JP-0140293.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI; 2000-039381/03.

PT New monoclonal antibodies, useful in diagnosis, as drugs and in
PT studying diseases related to ligand abnormality

PS Disclosure; Page 26; 73pp; Japanese.

CC The invention provides a monoclonal antibody which has a specific
CC reaction with the part peptide of the C-terminal of 19P2 ligand or its
CC derivative. The antibodies can be used in diagnosis or to treat or
CC prevent diseases associated with abnormality in the pituitary function
CC regulatory mechanism (e.g. promotion of prolactin secretion), central
CC nervous regulatory mechanism, and pancreatic function regulatory
CC mechanism. The antibody-based immunoassay can also be applied in
CC clarifying the physiological functions of the ligand and its derivative.

XX Human type G protein-coupled receptor ligand fragment 2.
 DE G protein-coupled receptor; ligand binding; pharmaceutical;
 KW modulator; pituitary; central nervous system; pancreas; prophylactic;
 KW therapeutic agent.
 XX
 OS Homo sapiens.
 XX
 PN WO9724436-A2.
 XX
 PD 10-JUL-1997.
 XX
 PF 26-DEC-1996; 96WO-JP03821.
 XX
 PR 18-SEP-1996; 96JP-0246573.
 XX 28-DEC-1995; 95JP-0343371.
 PR 15-MAR-1996; 96JP-0059419.
 PR 12-AUG-1996; 96JP-0211805.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
 PI Kawamata Y, Kltada C;
 XX
 DR WPI: 1997-363672/33.
 DR N-PSDB; AAV02429.
 XX
 PT Ligand peptide for G protein-coupled receptor - acts by modulating
 PT function in the central nervous system, pancreas and pituitary gland
 XX
 PS Claim 2; Page 185; 258pp; English.
 XX
 CC This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 23 to 54 of the
 CC sequence represented in AAW31390 and is used in an assay to monitor
 CC ligand binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a
 CC propylactic or therapeutic agent for dementia, depression, hyperkinetic
 CC syndrome, disturbance of consciousness, anxiety syndrome, hyperkinetic
 CC trauma, growth hormone secretory disease, hyper- and polyphagia,
 CC hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
 CC hyperprolactaemia, diabetes, cancer, pancreatitis, renal disease,
 CC Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
 CC transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
 CC acute myocardial infarction, infertility, spinocerebellar degeneration,
 CC bone fracture, trauma, atopic dermatitis, osteoporosis and/or
 CC oligogalactia. Assays can also be developed to screen compounds which are
 CC capable of altering the binding activity of the ligand affecting
 CC activation of the G protein-coupled receptor protein.
 CC
 XX
 SQ Sequence 32 AA;
 Query Match 100.0%; Score 171; DB 18; Length 32;
 Best Local Similarity 100.0%; Pred. No. 5.1e-19;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRTHRHSMETRPDINPAMVYASRGIRPVGRF 31
 DB 1 SRTHRHSMETRPDINPAMVYASRGIRPVGRF 31
 RESULT 9
 AAB10363 standard; peptide; 32 AA.
 AC AAB10363;
 XX
 XX 24-NOV-2000 (first entry)
 DT Human oxytocin secretion promoting peptide SEQ ID NO: 33.
 DE

XX Human: oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.
 XX
 OS Homo sapiens.
 XX
 PN WO200038704-A1.
 XX
 PD 06-JUL-2000.
 XX
 PF 22-DEC-1999; 99WO-JP07199.
 XX
 PR 25-DEC-1998; 98JP-0369585.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Matsumoto H, Kltada C, Hinuma S;
 XX
 DR WPI: 2000-452298/39.
 XX
 PT Physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine
 XX
 PS Disclosure; Page 62; 72pp; Japanese.
 XX
 CC This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,
 CC preventing and treating diseases relating to oxytocin secretion e.g.
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a human peptide which acts as an oxytocin secretion
 CC promoter.
 CC
 XX
 SQ Sequence 32 AA;
 Query Match 100.0%; Score 171; DB 21; Length 32;
 Best Local Similarity 100.0%; Pred. No. 5.1e-19;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SRTHRHSMETRPDINPAMVYASRGIRPVGRF 31
 DB 1 SRTHRHSMETRPDINPAMVYASRGIRPVGRF 31
 RESULT 10
 AAG62532 standard; peptide; 32 AA.
 ID AAG62532;
 XX
 XX AAG62532;
 AC
 XX
 DT 24-AUG-2001 (first entry)
 XX
 DE Human CRH releasing protein related peptide SEQ ID NO: 33.
 KW Human: corticotrophin releasing hormone; CRH; G protein receptor ligand;
 KW analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;
 KW Addison's disease; adrenal gland hyperfunction; obesity.
 XX
 OS Homo sapiens.
 XX
 PN WO200135984-A1.
 XX
 PD 25-MAY-2001.
 XX
 PF 17-NOV-2000; 2000WO-JP08119.
 XX

PR 18-NOV-1999; 99JP-0327900.
PR 26-SEP-2000; 2000JP-0297073.
XX (TAKE) TAKEDA CHEM IND LTD.
XX Kitada C, Matsumoto H, Hinuma S;
XX WPI; 2001-355552/37.
XX Use of G protein receptor ligand or peptide for controlling
XX corticotropin releasing hormone secretion .
XX Disclosure; Page 74; 90pp; Japanese.
XX The present sequence describes a method of controlling the secretion of
XX corticotropin releasing hormone (CRH), involving the use of a G protein
XX receptor ligand. This can be used to control the secretion of CRH and is
XX useful as an analgesic or for treating, preventing or ameliorating
XX diseases associated with CRH secretion such as hyperaldosteronism,
XX hypercortisolemia, secondary or chronic hypoadrenocorticism, Addison's
XX disease (including boredom, nausea, pigmentation, hypogonadism, hair
XX loss, and hypotension), adrenal gland hypofunction and obesity. The
XX present sequence is a peptide used in the exemplification of the
XX invention.
XX Sequence 32 AA:
SQ
Query Match 100.0%; Score 171; DB 22; Length 32;
Best Local Similarity 100.0%; Pred. No. 5.1e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SRTHRHSMETRTPDINPAMYASRGIRPVGRF 31
DB 1 SRTHRHSMETRTPDINPAMYASRGIRPVGRF 31
RESULT 11
AAM31393
ID AAM31393 standard; Peptide; 33 AA.
XX
XX AAM31393;
XX
XX 06-APR-1998 (first entry)
XX
XX Human type G protein-coupled receptor ligand fragment 3.
XX
XX G protein-coupled receptor; ligand binding; pharmaceutical;
XX modulator; pituitary; central nervous system; pancreas; prophylactic;
XX therapeutic agent.
XX
XX Homo sapiens.
XX
XX WO9724436-A2.
XX
XX 10-JUL-1997.
XX
XX 26-DEC-1996; 96WO-JP03821.
XX
XX 18-SEP-1996; 96JP-0246573.
XX 28-DEC-1995; 95JP-0343371.
XX 15-MAR-1996; 96JP-0059419.
XX 12-ADG-1996; 96JP-0211805.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Fuji R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
XX Kawamata Y, Kitada C;
XX WPI; 1997-363672/33.
XX N-PSDB; AAV02430.
XX
XX Ligand peptide for G protein-coupled receptor - acts by modulating
XX function in the central nervous system, pancreas and pituitary gland
XX

XX Claim 2; Page 185; 258pp; English.
XX
XX This sequence represents a peptide fragment from a novel human type
XX ligand polypeptide corresponding to amino acid residues 23 to 55 of the
XX sequence represented in AAM31390 and is used in an assay to monitor
XX ligand binding to the G protein-coupled receptor protein. Pharmaceutical
XX compositions containing this ligand may be used as a pituitary function
XX modulator, a central nervous system modulator or a pancreatic function
XX modulator. This ligand could have specific applications as a
XX prophylactic or therapeutic agent for dementia, depression, hyperkinetic
XX syndrome, disturbance of consciousness, anxiety syndrome, schizophrenia,
XX trauma, growth hormone secretory disease, hyper- and polyphagia,
XX hyperlipidaemia, hypercholesterolaemia, hyperglycaemia,
XX Turner's syndrome, neurosis, asthma, rheumatoid arthritis, spinal injury,
XX transient brain ischaemia, epilepsy, amyotrophic lateral sclerosis,
XX acute myocardial infarction, infertility, spinocerebellar degeneration,
XX bone fracture, trauma, atopic dermatitis, osteoporosis and/or
XX oligosaccharia. Assays can also be developed to screen compounds which are
XX capable of altering the binding activity of the ligand affecting
XX activation of the G protein-coupled receptor protein.
XX
XX Sequence 33 AA:
SQ
Query Match 100.0%; Score 171; DB 18; Length 33;
Best Local Similarity 100.0%; Pred. No. 5.3e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SRTHRHSMETRTPDINPAMYASRGIRPVGRF 31
DB 1 SRTHRHSMETRTPDINPAMYASRGIRPVGRF 31
RESULT 12
AAB10364
ID AAB10364 standard; peptide; 33 AA.
XX
XX AAB10364;
XX
XX 24-NOV-2000 (first entry)
XX
XX Human oxytocin secretion promoting peptide SEQ ID NO: 34.
XX
XX Human; oxytocin secretion promoter; G protein-coupled receptor protein;
XX treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
XX caesarian section; artificial fertilization; galactostasis; goat; pig;
XX veterinary medicine; milk production.
XX
XX Homo sapiens.
XX
XX WO200038704-A1.
XX
XX 06-JUL-2000.
XX
XX 22-DEC-1999; 99WO-JP07199.
XX
XX 25-DEC-1998; 98JP-0369585.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Matsumoto H, Kitada C, Hinuma S;
XX WPI; 2000-452298/39.
XX
XX Physiologically-active polypeptide recognized as ligand by G
XX protein-coupled receptor protein, for promoting secretion of oxytocin,
XX as drugs for diseases relating to oxytocin secretion and in veterinary
XX medicine -
XX Disclosure; Page 62-63; 72pp; Japanese.
XX
XX This invention describes a novel oxytocin secretion-regulating agent
XX

CC which contains a ligand peptide or its salt for the G protein-coupled
CC receptor protein. It is useful in the form of drugs for ameliorating,
CC preventing and treating diseases relating to oxytocin secretion e.g.
CC weak pains and atonic bleeding, before and after expulsion of placenta,
CC uterine recovery failure, caesarean section, stoppage of artificial
CC fertilization or galactostasis and is also applicable in veterinary
CC medicine for promoting milk production in cow, goat and pig. This
CC sequence represents a human peptide which acts as an oxytocin secretion
CC promoter.
XX

SQ Sequence 33 AA;

Query Match Best Local Similarity 100.0%; Score 171; DB 21; Length 33;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTPDINPAMVASRGIRPVGRF 31
DB 1 SRTHRSMETRTPDINPAMVASRGIRPVGRF 31

RESULT 13
AAG62533 ID AAG62533 standard; peptide; 33 AA.

XX AAG62533;

XX 24-AUG-2001 (first entry)

XX Human CRH releasing protein related peptide SEQ ID NO: 34.

XX Human; corticotrophin releasing hormone; CRH; G protein receptor ligand;
XX analgesic; hyperaldosteronism; hypercortisolemia; hypoadrenocorticism;
XX Addison's disease; adrenal gland hyperfunction; obesity.
XX Homo sapiens.

XX WO200135984-A1.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-JP08119.

XX 18-NOV-1999; 99JP-0327900.

XX 26-SEP-2000; 2000JP-0297073.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Kitada C, Matsumoto H, Hinuma S;

XX WPI; 2001-35552/37.

XX Use of G protein receptor ligand or peptide for controlling
XX corticotrophin releasing hormone secretion -
XX

XX Disclosure; Page 74; 90pp; Japanese.

XX The present sequence describes a method of controlling the secretion of
XX corticotrophin releasing hormone (CRH), involving the use of a G protein
XX receptor ligand. This can be used to control the secretion of CRH and is
XX useful as an analgesic or for treating, preventing or ameliorating
XX diseases associated with CRH secretion, such as hyperaldosteronism,
XX hypercortisolemia, secondary or chronic hypoadrenocorticism, Addison's
XX disease (including boredom, nausea, pigmentation, hypogonadism, hair
XX loss, and hypotension), adrenal gland hypofunction and obesity. The
XX present sequence is a peptide used in the exemplification of the
XX invention.
XX

SQ Sequence 33 AA;

Query Match Best Local Similarity 100.0%; Score 171; DB 22; Length 33;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTPDINPAMVASRGIRPVGRF 31
DB 1 SRTHRSMETRTPDINPAMVASRGIRPVGRF 31

RESULT 14
AAM31390 ID AAM31390 standard; protein; 87 AA.

XX AAM31390;

XX 06-APR-1998 (first entry)

XX Human type G protein-coupled receptor ligand encoded by PHO7.

XX G protein-coupled receptor; ligand binding; pharmaceutical;
XX modulator; pituitary; central nervous system; pancreas; prophylactic;
XX therapeutic agent.
XX

XX Homo sapiens.

XX WO9724436-A2.

XX 10-JUL-1997.

XX 26-DEC-1996; 96WO-JP03821.

XX 18-SEP-1996; 96JP-0246573.

XX 28-DEC-1995; 95JP-0343371.

XX 15-MAR-1996; 96JP-0059419.

XX 12-AUG-1996; 96JP-0211805.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;

XX Kawamata Y, Kitada C;

XX WPI; 1997-363672/33.

XX N-PSDB; AAV02427.

XX Ligand peptide for G protein-coupled receptor - acts by modulating
XX function in the central nervous system, pancreas and pituitary gland
XX

XX Claim 3; Page 183; 258pp; English.

XX This sequence represents a novel human type ligand polypeptide encoded
XX by PHO7 which is used in an assay to monitor ligand binding to the G
XX protein-coupled receptor protein. Pharmaceutical compositions
XX containing this ligand may be used as a pituitary function modulator, a
XX central nervous system modulator or a pancreatic function modulator. This
XX ligand could have specific applications as a prophylactic or therapeutic
XX agent for dementia, depression, hyperkinetic syndrome, disturbance of
XX consciousness, anxiety syndrome, schizophrenia, trauma, growth hormone
XX secretory disease, hyperlipidaemia, hypercholesterolemia,
XX hyperglycaemia, hyperlipidaemia, hyperproliferative disease,
XX cancer, pancreatitis, renal disease, Turner's syndrome, neurosis,
XX rheumatoid arthritis, spinal injury, transient brain ischaemia,
XX amyotrophic lateral sclerosis, acute myocardial infarction,
XX splintered bone degeneration, bone fracture, trauma, atopic dermatitis,
XX osteoporosis, asthma, epilepsy, infertility and/or oligogalactia. Assays
XX can also be developed to screen compounds which are capable of altering
XX the binding activity of the ligand thus affecting activation of the G
XX protein-coupled receptor protein.
XX

SQ Sequence 87 AA;

Query Match Best Local Similarity 100.0%; Score 171; DB 18; Length 87;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 23 SRTHRHSMETPTDINPAMYASRGIRPVGRF 53

RESULT 15

AAW97226 standard; peptide; 87 AA.

AAW97226;

06-MAY-1999 (first entry)

Human type ligand polypeptide.

G protein-coupled receptor; GPCR; hypovarianism; gonocyst cacogenesis;
menopausal syndrome; euthyroid; hypometabolism; lactation; modulation;
pituitary adenomatosis; brain tumour; emmenopathy; autoimmune disease;
prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
contraceptive; placental function; choriocarcinoma; hydatid mole;
irruption mole; abortion; unthrifty fetus; abnormal saccharometabolism;
human type ligand.

Homo sapiens.

MO9858962-A1.

30-DEC-1998.

22-JUN-1998; 98WC-JP02765.

23-JUN-1997; 97JP-0165437.

(TAKE) TAKEDA CHEM IND LTD.

Fuji R, Hinuma S, Kawamata Y, Matsumoto H;

WPI: 1999-105614/09.

N-PSDB: AAX1526.

Use of G protein-coupled receptor ligands - for modulating prolactin
secretion or placental function, e.g. for treating menopausal
syndrome, tumours, autoimmune disease or abnormal pregnancy

Disclosure: Page 158; 241pp; English.

The present sequence represents a human type ligand polypeptide. The
specification describes an agent for modulating prolactin secretion
which comprises a ligand polypeptide or a salt, for a G protein-coupled
receptor (GPCR) protein. The agents for promoting prolactin secretion
can be used for treating or preventing hypovarianism, gonocyst
cacogenesis, menopausal syndrome, euthyroid or hypometabolism. They
can be used for promoting lactation in a domestic mammal and as an
aphrodisiac. The agents for inhibiting prolactin secretion can be used
for treating or preventing pituitary adenomatosis, brain tumour,
amenorrhoea, galactorrhea, acromegaly, Chiari-Frommel syndrome, impotence,
amenorrhoea, autoimmune disease, prolactinoma, infertility, impotence,
Castillo syndrome, Forbes-Albright syndrome, lymphoma, Sheehan syndrome
or dyszoospermia. The inhibitory agents can also be used as
contraceptives. The agents for modulating placental function can be used
for treating or preventing choriocarcinoma, hydatid mole, irruption mole,
abortion, unthrifty fetus, abnormal saccharometabolism, abnormal
lipidmetabolism or oxytocia.

Sequence 87 AA;

Query Match

100.0%; Score 171; DB 20; Length 87;

Best Local Similarity 100.0%; Pred. No. 1.7e-18;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 SRTHRHSMETPTDINPAMYASRGIRPVGRF 31

DB 23 SRTHRHSMETPTDINPAMYASRGIRPVGRF 53

Search completed: May 1, 2003, 14:30:49

Job time : 44.1094 secs

Thu May 1 15:54:25 2003

us-09-446-543a-73.rsp

SEQ ID NO: 73
AC NO: P81277
Database: SwissProt_40

Page 1

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 7.21875 Seconds

(without alignments)
126.404 Million cell updates/sec

Title: US-09-446-543a-73

Perfect score: 109

Sequence: 1 TPDINPAMYXXRCIRPYGRFX 22

Scoring table: Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	96.3	87	1 PRRP_HUMAN	P81277 homo sapien
2	104	95.4	83	1 PRRP_RAT	P81278 rattus norv
3	104	95.4	98	1 PRRP_BOVIN	P81264 bos taurus
4	46	42.2	676	1 EXL1_HUMAN	O92935 homo sapien
5	43	39.4	376	1 OPS1_LIMPO	P33360 limulus pol
6	43	39.4	376	1 OPS2_LIMPO	P33361 limulus pol
7	43	39.4	383	1 CYCR_CHRVI	O82947 chromatium
8	43	39.4	476	1 YAAJ_ECOLI	P30143 escherichia
9	43	39.4	719	1 NRPL_YEAST	P32770 saccharomyc
10	42	38.5	402	1 EXTL_SYRGO	O9f8m3 streptomyce
11	41.5	38.1	345	1 ARGC_BACHD	O9f8v2 bacillus ha
12	41	37.6	342	1 Y762_METUA	O58172 methanococc
13	41	37.6	347	1 Y576_METUA	O57996 methanococc
14	41	37.6	986	1 CYGR_ARBPV	P11528 arabacia pun
15	40	36.7	329	1 Y493_MYCTU	O11158 mycobacteri
16	40	36.7	546	1 CHOD_STRSO	P12676 streptomyce
17	40	36.7	581	1 POL_MLYRK	P31795 streptomyce
18	40	36.7	697	1 ICFE_HUMAN	O9f8u0 homo sapien
19	40	36.7	843	1 POL_MLYRK	O9f8s2 ratelonia s
20	40	36.7	1024	1 POPC_RALSO	P03356 akr murine
21	40	36.7	1196	1 POL_MLYRK	P11227 radiation m
22	39.5	36.2	860	1 V612_BPB03	O37893 bacterioph
23	39.5	35.8	149	1 ENRN_BPT7	P00641 bacterioph
24	39	35.8	360	1 WNR2_HUMAN	O09544 homo sapien
25	39	35.8	377	1 OPS1_HEMSA	O25157 hemigrapsus
26	39	35.8	377	1 OPS2_HEMSA	O25158 hemigrapsus
27	39	35.8	485	1 SYE_BACHD	O9f8f6 bacillus ha
28	39	35.8	622	1 PCC_RAT	P07378 rattus norv
29	39	35.8	953	1 SYV_VIBCH	O9f8p7 vibrato chol
30	39	35.8	962	1 UYVA_METTH	O28543 methanobact
31	39	35.8	962	1 UYVA_METTH	O28543 methanobact
32	38.5	35.3	240	1 PLSC_HELRY	O92905 helicobacte
33	38.5	35.3	248	1 UBIE_RICCN	O929t5 rickettsia

ALIGNMENTS

ID	PRRP_HUMAN	STANDARD	PRT	87 AA
AC	P81277			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Proactin-releasing peptide precursor (PRPP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PRP31; Proactin-releasing peptide PRP20].			
DE	releasing peptide PRP20].			
GN	PRP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=98268781; PubMed=9607765;			
RA	Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kikada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M., Nature 393:272-276(1998).			
RT	"A prolactin-releasing peptide in the brain."			
RL	Nature 393:272-276(1998).			
RN	[2]			
RP	TISSUE SPECIFICITY.			
RX	MEDLINE=9942652; PubMed=10498338;			
RA	Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kikada C., Kurokawa T., Nishimura O., Onda H., Sumino Y., Fujino M., "Tissue distribution of prolactin-releasing peptide (PRP) and its receptor."			
RT	Regul. Pept. 83:1-10(1999).			
RL	-1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.			
CC	-1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).			
CC	EMBL, AB015419; BAA29027.1; -			
DR	MIM: 602663; -			
KW	Hormone; Amidation; Signal.			
FT	STANDARD			
FT	1			
FT	PEPTIDE			
FT	PEPTIDE			
FT	MOD. RES			
FT	53			
FT	SEQUENCE			
FT	87 AA; 9639 MW;			
FT	229A2F3F50CF981B CRC64;			

Query Match 96.3%; Score 105; DB 1; Length 87;

Best Local Similarity 90.0%; Pred. No. 1,1e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||||
Db 34 TPDINPAMYXXRGIRPVGRF 53

RESULT 2

PRRP_RAT STANDARD; PRT; 83 AA.

AC P81278;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PrRP) (Proactin-releasing hormone) [contains: Proactin-releasing peptide PrRP31; Proactin-releasing peptide PrRP20].

GN PRH.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_TaxID=10116;

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA MEDLINE=98268781; PubMed=9607765;

RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,

RA Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,

RA Kurokawa T., Nishimura O., Onda H., Fujino M.;

RT "A prolactin-releasing peptide in the brain.";

RL Nature 393:272-276(1998).

CC [2]

CC TISSUE SPECIFICITY:

CC MEDLINE=9942652; PubMed=10498338;

RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Hinuma S.,

RA Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H.,

RA Sumino Y., Fujino M.;

RT "Tissue distribution of prolactin-releasing peptide (PrRP) and its

RT receptor.";

RL Regul. Pept. 83:1-10(1999).

CC -1- FUNCTION: Stimulates prolactin (PR) release and regulates the

CC expression of prolactin through its receptor GPR10. May stimulate

CC lactotrophs directly to secrete PR.

CC -1- TISSUE SPECIFICITY: Widely expressed, with highest levels in

CC medulla oblongata and hypothalamus.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AB015418; BAA29026.1;

KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.

FT SIGNAL 1 21 BY SIMILARITY.

FT PEPTIDE 22 52 PROLACTIN-RELEASING PEPTIDE PRRP31.

FT PEPTIDE 33 52 PROLACTIN-RELEASING PEPTIDE PRRP20.

FT MOD RES 52 52 AMIDATION (G-53 PROVIDE AMIDE GROUP).

SO SEQUENCE 83 AA: 9215 MW; DDCG5A264EEAF29 CRC64;

Query Match 95.4%; Score 104; DB 1; Length 83;

Best Local Similarity 90.0%; Pred. No. 1.6e-10;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20

|||||||

Db 33 TPDINPAMYXXRGIRPVGRF 52

RESULT 3

PRRP_BOVIN STANDARD; PRT; 98 AA.

AC P81264;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Proactin-releasing peptide precursor (PrRP) (Proactin-releasing hormone) [contains: Proactin-releasing peptide PrRP31; Proactin-releasing peptide PrRP20].

GN PRH.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RP SEQUENCE FROM N.A. AND SEQUENCE OF 23-52.

RC TISSUE=Brain;

RA MEDLINE=98268781; PubMed=9607765;

RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,

RA Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,

RA Kurokawa T., Nishimura O., Onda H., Fujino M.;

RT "A prolactin-releasing peptide in the brain.";

RL Nature 393:272-276(1998).

CC -1- FUNCTION: Stimulates prolactin (PR) release and regulates the

CC expression of prolactin through its receptor GPR10. May stimulate

CC lactotrophs directly to secrete PR.

CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AB015417; BAA29025.1;

KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.

FT SIGNAL 1 22 BY SIMILARITY.

FT PEPTIDE 23 53 PROLACTIN-RELEASING PEPTIDE PRRP31.

FT PEPTIDE 33 53 PROLACTIN-RELEASING PEPTIDE PRRP20.

FT MOD RES 53 53 AMIDATION (G-54 PROVIDE AMIDE GROUP).

SO SEQUENCE 98 AA: 10544 MW; 08AC35A13B0FA908 CRC64;

Query Match 95.4%; Score 104; DB 1; Length 98;

Best Local Similarity 90.0%; Pred. No. 1.8e-10;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20

|||||||

Db 34 TPDINPAMYXXRGIRPVGRF 53

RESULT 4

EX1L_HUMAN STANDARD; PRT; 676 AA.

AC Q92935;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Exostosin-like 1 (EC 2.4.1.-) (Exostosin-L) (Multiple exostosin-like protein).

GN EX1L OR EXTL.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RA MEDLINE=97189339; PubMed=9037597;

RT "Identification and localization of the gene for EX1L, a third member

RT of the multiple exostoses gene family.";
RN Genome Res. 7:10-16(1997).
[2]
RP SEQUENCE FROM N.A. AND VARIANT HIS-379.
RX MEDLINE-99408231; PubMed-10480354;
RA Xu L., Xia J., Jiang H., Zhou J., Li H., Wang D., Pan Q., Long Z.,
R Fan C., Deng H.-X.,
RT "Mutation analysis of hereditary multiple exostoses in the Chinese.";
RN Hum. Genet. 105:45-50(1999).
[3]
RP SEQUENCE FROM N.A.
RA Wuyts W., Spleiter N., Van Roy N., De Paeppe A., De Boule K.,
RA Williams P.J., Van Hul W., Versteeg R., Speleman F.,
RT "Refined physical mapping and genomic structure of the EXT1 gene.";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probable glycosyltransferase (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 47.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67191; AAC51141.1; -
DR EMBL: AF083633; AAD02840.1; -
DR EMBL: AF083623; AAD02840.1; JOINED.
DR EMBL: AF083624; AAD02840.1; JOINED.
DR EMBL: AF083625; AAD02840.1; JOINED.
DR EMBL: AF083626; AAD02840.1; JOINED.
DR EMBL: AF083627; AAD02840.1; JOINED.
DR EMBL: AF083628; AAD02840.1; JOINED.
DR EMBL: AF083629; AAD02840.1; JOINED.
DR EMBL: AF083630; AAD02840.1; JOINED.
DR EMBL: AF083631; AAD02840.1; JOINED.
DR EMBL: AF083632; AAD02840.1; JOINED.
DR EMBL: AF153980; AAF73172.1; -
DR EMBL: AF153981; AAF73172.1; JOINED.
DR Gene: HGNC:3515; EXT1.
DR MIM: 601738; -
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin.1.
KW Transferase; Glycosyltransferase; Endoplasmic reticulum;
KW Transmembrane; Signal-anchor; Glycoprotein; Polymorphism.
FT TRANSMEM 1 9
FT DOMAIN 1 30
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT LOMENAL (POTENTIAL).
FT CARBOHYD 269 269
FT VARIANT 379 379
FT N->H.
FT /FTID=VAR_012830.
SQ SEQUENCE 676 AA; 74673 MW; B5E006A8762E5633 CRC64;
Query Match 42.2%; Score 46; DB 1; Length 676;
Best Local Similarity 45.0%; Pred. No. 4.9;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 400 SPDDPFYIQQSGRPEGGRF 419
RESULT 5
OPSL LIMPO STANDARD; PRT; 376 AA.
AC P35360;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Lateral eye opsin.
OS Limulus polyphemus (Atlantic horseshoe crab).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC Limulidae; Limulus.
OX NCBI-Taxid=6850;
RN (1)
RP SEQUENCE FROM N.A.
RA Smith W.C., Price D.A., Greenberg R.M., Battelle B.-A.;
RX MEDLINE-93317641; PubMed-8327495;
RT "Opsins from the lateral eyes and ocell of the horseshoe crab,
RT Limulus polyphemus.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6150-6154(1993).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: LATERAL EYE.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED (BY SIMILARITY).
CC -1- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMUM AT 520 NM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L03791; AAA28273.1; -
DR EMBL: L03781; AAA02498.1; -
DR PIR: B48197; B48197.
DR HSPR: P02699; 1EDV.
DR InterPro: IPR00276; GPCR_Rhodopsn.
DR InterPro: IPR001760; Opsin.
DR Pfam: PF00001; 7tm_1.1.
DR PRINTS: PR00237; GPCR_Rhodopsn.
DR PROSITE: PS00237; G-PROTEIN_RECPT_FL_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECPT_FL_2; 1.
DR PROSITE: PS00238; OPSIN.1.
DR KMW Photorceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW Phosphorylation; G-protein coupled receptor.
FT DOMAIN 1 46
FT TRANSMEM 47 71
FT DOMAIN 72 83
FT TRANSMEM 84 108
FT DOMAIN 109 123
FT TRANSMEM 124 143
FT DOMAIN 144 162
FT TRANSMEM 163 186
FT DOMAIN 187 210
FT TRANSMEM 211 238
FT DOMAIN 239 274
FT TRANSMEM 275 298
FT DOMAIN 299 306
FT TRANSMEM 307 331
FT DOMAIN 332 376
FT DISULF 120 197
FT BINDING 318 318
FT CARBOHYD 17 17
FT FT 193 193
SQ SEQUENCE 376 AA; 42139 MW; CCE401766AB06F26 CRC64;
Query Match 39.4%; Score 43; DB 1; Length 376;
Best Local Similarity 46.2%; Pred. No. 8.7;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 2 PDINPAMYXXRGI 14
DB 40 PPMNLMWYSILGV 52

RESULT 6
 OPS2_LIMPO STANDARD; PRT; 376 AA.
 AC P35361;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ocellar opsin.
 OS Limulus polyphemus (Atlantic horseshoe crab).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
 OC Limulidae; Limulus.
 OX NCBI_TaxID=6850;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Median ocellar;
 RX MEDLINE=93317641; PubMed=8327495;
 RA Smith W.C., Price D.A., Greenberg R.M., Battelle B.A.;
 RT "Opsins from the lateral eyes and ocelli of the horseshoe crab,
 Limulus polyphemus."
 RT Proc. Natl. Acad. Sci. U.S.A. 90:6150-6154(1993).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: OCELLAR CELLS; MEDIAN OCELLI.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 BE PHOSPHORYLATED (BY SIMILARITY).
 CC -1- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMUM AT 530 NM.
 CC -1- SIMILARITY: BEYONDS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L03792; AAA28274.1; -;
 CC EMBL; L03782; AAA02499.1; -;
 CC PIR; A48197; A48197.
 CC HSSP; P02699; LEDV.
 CC InterPro: IPR000276; GPCR_Rhodopsin.
 CC InterPro: IPR001760; Opsin.
 CC Pfam: PF00001; 7tm.1; 1.
 CC PRINTS: PR00237; GPCRHOPOPSN.
 CC PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
 CC PROSITE; PS50262; G-PROTEIN_RECEP_F1_2; 1.
 CC PROSITE; PS00238; OPSIN; 1.
 CC PhotoReceptor: Retinal protein; Transmembrane; Glycoprotein; Vision;
 KW Phosphorylation; G-protein coupled receptor.
 FT DOMAIN 1 46
 FT TRANSMEM 47 71
 FT TRANSMEM 72 83
 FT TRANSMEM 84 108
 FT TRANSMEM 109 123
 FT TRANSMEM 124 143
 FT TRANSMEM 144 162
 FT TRANSMEM 163 186
 FT TRANSMEM 187 210
 FT TRANSMEM 211 238
 FT TRANSMEM 239 274
 FT TRANSMEM 275 298
 FT TRANSMEM 299 306
 FT TRANSMEM 307 331
 FT TRANSMEM 332 376
 FT TRANSMEM 377 397
 FT TRANSMEM 398 418
 FT TRANSMEM 419 438
 FT TRANSMEM 439 459
 FT TRANSMEM 460 480
 FT TRANSMEM 481 500
 FT TRANSMEM 501 520
 FT TRANSMEM 521 540
 FT TRANSMEM 541 560
 FT TRANSMEM 561 580
 FT TRANSMEM 581 600
 FT TRANSMEM 601 620
 FT TRANSMEM 621 640
 FT TRANSMEM 641 660
 FT TRANSMEM 661 680
 FT TRANSMEM 681 700
 FT TRANSMEM 701 720
 FT TRANSMEM 721 740
 FT TRANSMEM 741 760
 FT TRANSMEM 761 780
 FT TRANSMEM 781 800
 FT TRANSMEM 801 820
 FT TRANSMEM 821 840
 FT TRANSMEM 841 860
 FT TRANSMEM 861 880
 FT TRANSMEM 881 900
 FT TRANSMEM 901 920
 FT TRANSMEM 921 940
 FT TRANSMEM 941 960
 FT TRANSMEM 961 980
 FT TRANSMEM 981 1000
 FT TRANSMEM 1001 1020
 FT TRANSMEM 1021 1040
 FT TRANSMEM 1041 1060
 FT TRANSMEM 1061 1080
 FT TRANSMEM 1081 1100
 FT TRANSMEM 1101 1120
 FT TRANSMEM 1121 1140
 FT TRANSMEM 1141 1160
 FT TRANSMEM 1161 1180
 FT TRANSMEM 1181 1200
 FT TRANSMEM 1201 1220
 FT TRANSMEM 1221 1240
 FT TRANSMEM 1241 1260
 FT TRANSMEM 1261 1280
 FT TRANSMEM 1281 1300
 FT TRANSMEM 1301 1320
 FT TRANSMEM 1321 1340
 FT TRANSMEM 1341 1360
 FT TRANSMEM 1361 1380
 FT TRANSMEM 1381 1400
 FT TRANSMEM 1401 1420
 FT TRANSMEM 1421 1440
 FT TRANSMEM 1441 1460
 FT TRANSMEM 1461 1480
 FT TRANSMEM 1481 1500
 FT TRANSMEM 1501 1520
 FT TRANSMEM 1521 1540
 FT TRANSMEM 1541 1560
 FT TRANSMEM 1561 1580
 FT TRANSMEM 1581 1600
 FT TRANSMEM 1601 1620
 FT TRANSMEM 1621 1640
 FT TRANSMEM 1641 1660
 FT TRANSMEM 1661 1680
 FT TRANSMEM 1681 1700
 FT TRANSMEM 1701 1720
 FT TRANSMEM 1721 1740
 FT TRANSMEM 1741 1760
 FT TRANSMEM 1761 1780
 FT TRANSMEM 1781 1800
 FT TRANSMEM 1801 1820
 FT TRANSMEM 1821 1840
 FT TRANSMEM 1841 1860
 FT TRANSMEM 1861 1880
 FT TRANSMEM 1881 1900
 FT TRANSMEM 1901 1920
 FT TRANSMEM 1921 1940
 FT TRANSMEM 1941 1960
 FT TRANSMEM 1961 1980
 FT TRANSMEM 1981 2000
 FT TRANSMEM 2001 2020
 FT TRANSMEM 2021 2040
 FT TRANSMEM 2041 2060
 FT TRANSMEM 2061 2080
 FT TRANSMEM 2081 2100
 FT TRANSMEM 2101 2120
 FT TRANSMEM 2121 2140
 FT TRANSMEM 2141 2160
 FT TRANSMEM 2161 2180
 FT TRANSMEM 2181 2200
 FT TRANSMEM 2201 2220
 FT TRANSMEM 2221 2240
 FT TRANSMEM 2241 2260
 FT TRANSMEM 2261 2280
 FT TRANSMEM 2281 2300
 FT TRANSMEM 2301 2320
 FT TRANSMEM 2321 2340
 FT TRANSMEM 2341 2360
 FT TRANSMEM 2361 2380
 FT TRANSMEM 2381 2400
 FT TRANSMEM 2401 2420
 FT TRANSMEM 2421 2440
 FT TRANSMEM 2441 2460
 FT TRANSMEM 2461 2480
 FT TRANSMEM 2481 2500
 FT TRANSMEM 2501 2520
 FT TRANSMEM 2521 2540
 FT TRANSMEM 2541 2560
 FT TRANSMEM 2561 2580
 FT TRANSMEM 2581 2600
 FT TRANSMEM 2601 2620
 FT TRANSMEM 2621 2640
 FT TRANSMEM 2641 2660
 FT TRANSMEM 2661 2680
 FT TRANSMEM 2681 2700
 FT TRANSMEM 2701 2720
 FT TRANSMEM 2721 2740
 FT TRANSMEM 2741 2760
 FT TRANSMEM 2761 2780
 FT TRANSMEM 2781 2800
 FT TRANSMEM 2801 2820
 FT TRANSMEM 2821 2840
 FT TRANSMEM 2841 2860
 FT TRANSMEM 2861 2880
 FT TRANSMEM 2881 2900
 FT TRANSMEM 2901 2920
 FT TRANSMEM 2921 2940
 FT TRANSMEM 2941 2960
 FT TRANSMEM 2961 2980
 FT TRANSMEM 2981 3000
 FT TRANSMEM 3001 3020
 FT TRANSMEM 3021 3040
 FT TRANSMEM 3041 3060
 FT TRANSMEM 3061 3080
 FT TRANSMEM 3081 3100
 FT TRANSMEM 3101 3120
 FT TRANSMEM 3121 3140
 FT TRANSMEM 3141 3160
 FT TRANSMEM 3161 3180
 FT TRANSMEM 3181 3200
 FT TRANSMEM 3201 3220
 FT TRANSMEM 3221 3240
 FT TRANSMEM 3241 3260
 FT TRANSMEM 3261 3280
 FT TRANSMEM 3281 3300
 FT TRANSMEM 3301 3320
 FT TRANSMEM 3321 3340
 FT TRANSMEM 3341 3360
 FT TRANSMEM 3361 3380
 FT TRANSMEM 3381 3400
 FT TRANSMEM 3401 3420
 FT TRANSMEM 3421 3440
 FT TRANSMEM 3441 3460
 FT TRANSMEM 3461 3480
 FT TRANSMEM 3481 3500
 FT TRANSMEM 3501 3520
 FT TRANSMEM 3521 3540
 FT TRANSMEM 3541 3560
 FT TRANSMEM 3561 3580
 FT TRANSMEM 3581 3600
 FT TRANSMEM 3601 3620
 FT TRANSMEM 3621 3640
 FT TRANSMEM 3641 3660
 FT TRANSMEM 3661 3680
 FT TRANSMEM 3681 3700
 FT TRANSMEM 3701 3720
 FT TRANSMEM 3721 3740
 FT TRANSMEM 3741 3760
 FT TRANSMEM 3761 3780
 FT TRANSMEM 3781 3800
 FT TRANSMEM 3801 3820
 FT TRANSMEM 3821 3840
 FT TRANSMEM 3841 3860
 FT TRANSMEM 3861 3880
 FT TRANSMEM 3881 3900
 FT TRANSMEM 3901 3920
 FT TRANSMEM 3921 3940
 FT TRANSMEM 3941 3960
 FT TRANSMEM 3961 3980
 FT TRANSMEM 3981 4000
 FT TRANSMEM 4001 4020
 FT TRANSMEM 4021 4040
 FT TRANSMEM 4041 4060
 FT TRANSMEM 4061 4080
 FT TRANSMEM 4081 4100
 FT TRANSMEM 4101 4120
 FT TRANSMEM 4121 4140
 FT TRANSMEM 4141 4160
 FT TRANSMEM 4161 4180
 FT TRANSMEM 4181 4200
 FT TRANSMEM 4201 4220
 FT TRANSMEM 4221 4240
 FT TRANSMEM 4241 4260
 FT TRANSMEM 4261 4280
 FT TRANSMEM 4281 4300
 FT TRANSMEM 4301 4320
 FT TRANSMEM 4321 4340
 FT TRANSMEM 4341 4360
 FT TRANSMEM 4361 4380
 FT TRANSMEM 4381 4400
 FT TRANSMEM 4401 4420
 FT TRANSMEM 4421 4440
 FT TRANSMEM 4441 4460
 FT TRANSMEM 4461 4480
 FT TRANSMEM 4481 4500
 FT TRANSMEM 4501 4520
 FT TRANSMEM 4521 4540
 FT TRANSMEM 4541 4560
 FT TRANSMEM 4561 4580
 FT TRANSMEM 4581 4600
 FT TRANSMEM 4601 4620
 FT TRANSMEM 4621 4640
 FT TRANSMEM 4641 4660
 FT TRANSMEM 4661 4680
 FT TRANSMEM 4681 4700
 FT TRANSMEM 4701 4720
 FT TRANSMEM 4721 4740
 FT TRANSMEM 4741 4760
 FT TRANSMEM 4761 4780
 FT TRANSMEM 4781 4800
 FT TRANSMEM 4801 4820
 FT TRANSMEM 4821 4840
 FT TRANSMEM 4841 4860
 FT TRANSMEM 4861 4880
 FT TRANSMEM 4881 4900
 FT TRANSMEM 4901 4920
 FT TRANSMEM 4921 4940
 FT TRANSMEM 4941 4960
 FT TRANSMEM 4961 4980
 FT TRANSMEM 4981 5000
 FT TRANSMEM 5001 5020
 FT TRANSMEM 5021 5040
 FT TRANSMEM 5041 5060
 FT TRANSMEM 5061 5080
 FT TRANSMEM 5081 5100
 FT TRANSMEM 5101 5120
 FT TRANSMEM 5121 5140
 FT TRANSMEM 5141 5160
 FT TRANSMEM 5161 5180
 FT TRANSMEM 5181 5200
 FT TRANSMEM 5201 5220
 FT TRANSMEM 5221 5240
 FT TRANSMEM 5241 5260
 FT TRANSMEM 5261 5280
 FT TRANSMEM 5281 5300
 FT TRANSMEM 5301 5320
 FT TRANSMEM 5321 5340
 FT TRANSMEM 5341 5360
 FT TRANSMEM 5361 5380
 FT TRANSMEM 5381 5400
 FT TRANSMEM 5401 5420
 FT TRANSMEM 5421 5440
 FT TRANSMEM 5441 5460
 FT TRANSMEM 5461 5480
 FT TRANSMEM 5481 5500
 FT TRANSMEM 5501 5520
 FT TRANSMEM 5521 5540
 FT TRANSMEM 5541 5560
 FT TRANSMEM 5561 5580
 FT TRANSMEM 5581 5600
 FT TRANSMEM 5601 5620
 FT TRANSMEM 5621 5640
 FT TRANSMEM 5641 5660
 FT TRANSMEM 5661 5680
 FT TRANSMEM 5681 5700
 FT TRANSMEM 5701 5720
 FT TRANSMEM 5721 5740
 FT TRANSMEM 5741 5760
 FT TRANSMEM 5761 5780
 FT TRANSMEM 5781 5800
 FT TRANSMEM 5801 5820
 FT TRANSMEM 5821 5840
 FT TRANSMEM 5841 5860
 FT TRANSMEM 5861 5880
 FT TRANSMEM 5881 5900
 FT TRANSMEM 5901 5920
 FT TRANSMEM 5921 5940
 FT TRANSMEM 5941 5960
 FT TRANSMEM 5961 5980
 FT TRANSMEM 5981 6000
 FT TRANSMEM 6001 6020
 FT TRANSMEM 6021 6040
 FT TRANSMEM 6041 6060
 FT TRANSMEM 6061 6080
 FT TRANSMEM 6081 6100
 FT TRANSMEM 6101 6120
 FT TRANSMEM 6121 6140
 FT TRANSMEM 6141 6160
 FT TRANSMEM 6161 6180
 FT TRANSMEM 6181 6200
 FT TRANSMEM 6201 6220
 FT TRANSMEM 6221 6240
 FT TRANSMEM 6241 6260
 FT TRANSMEM 6261 6280
 FT TRANSMEM 6281 6300
 FT TRANSMEM 6301 6320
 FT TRANSMEM 6321 6340
 FT TRANSMEM 6341 6360
 FT TRANSMEM 6361 6380
 FT TRANSMEM 6381 6400
 FT TRANSMEM 6401 6420
 FT TRANSMEM 6421 6440
 FT TRANSMEM 6441 6460
 FT TRANSMEM 6461 6480
 FT TRANSMEM 6481 6500
 FT TRANSMEM 6501 6520
 FT TRANSMEM 6521 6540
 FT TRANSMEM 6541 6560
 FT TRANSMEM 6561 6580
 FT TRANSMEM 6581 6600
 FT TRANSMEM 6601 6620
 FT TRANSMEM 6621 6640
 FT TRANSMEM 6641 6660
 FT TRANSMEM 6661 6680
 FT TRANSMEM 6681 6700
 FT TRANSMEM 6701 6720
 FT TRANSMEM 6721 6740
 FT TRANSMEM 6741 6760
 FT TRANSMEM 6761 6780
 FT TRANSMEM 6781 6800
 FT TRANSMEM 6801 6820
 FT TRANSMEM 6821 6840
 FT TRANSMEM 6841 6860
 FT TRANSMEM 6861 6880
 FT TRANSMEM 6881 6900
 FT TRANSMEM 6901 6920
 FT TRANSMEM 6921 6940
 FT TRANSMEM 6941 6960
 FT TRANSMEM 6961 6980
 FT TRANSMEM 6981 7000
 FT TRANSMEM 7001 7020
 FT TRANSMEM 7021 7040
 FT TRANSMEM 7041 7060
 FT TRANSMEM 7061 7080
 FT TRANSMEM 7081 7100
 FT TRANSMEM 7101 7120
 FT TRANSMEM 7121 7140
 FT TRANSMEM 7141 7160
 FT TRANSMEM 7161 7180
 FT TRANSMEM 7181 7200
 FT TRANSMEM 7201 7220
 FT TRANSMEM 7221 7240
 FT TRANSMEM 7241 7260
 FT TRANSMEM 7261 7280
 FT TRANSMEM 7281 7300
 FT TRANSMEM 7301 7320
 FT TRANSMEM 7321 7340
 FT TRANSMEM 7341 7360
 FT TRANSMEM 7361 7380
 FT TRANSMEM 7381 7400
 FT TRANSMEM 7401 7420
 FT TRANSMEM 7421 7440
 FT TRANSMEM 7441 7460
 FT TRANSMEM 7461 7480
 FT TRANSMEM 7481 7500
 FT TRANSMEM 7501 7520
 FT TRANSMEM 7521 7540
 FT TRANSMEM 7541 7560
 FT TRANSMEM 7561 7580
 FT TRANSMEM 7581 7600
 FT TRANSMEM 7601 7620
 FT TRANSMEM 7621 7640
 FT TRANSMEM 7641 7660
 FT TRANSMEM 7661 7680
 FT TRANSMEM 7681 7700
 FT TRANSMEM 7701 7720
 FT TRANSMEM 7721 7740
 FT TRANSMEM 7741 7760
 FT TRANSMEM 7761 7780
 FT TRANSMEM 7781 7800
 FT TRANSMEM 7801 7820
 FT TRANSMEM 7821 7840
 FT TRANSMEM 7841 7860
 FT TRANSMEM 7861 7880
 FT TRANSMEM 7881 7900
 FT TRANSMEM 7901 7920
 FT TRANSMEM 7921 7940
 FT TRANSMEM 7941 7960
 FT TRANSMEM 7961 7980
 FT TRANSMEM 7981 8000
 FT TRANSMEM 8001 8020
 FT TRANSMEM 8021 8040
 FT TRANSMEM 8041 8060
 FT TRANSMEM 8061 8080
 FT TRANSMEM 8081 8100
 FT TRANSMEM 8101 8120
 FT TRANSMEM 8121 8140
 FT TRANSMEM 8141 8160
 FT TRANSMEM 8161 8180
 FT TRANSMEM 8181 8200
 FT TRANSMEM 8201 8220
 FT TRANSMEM 8221 8240
 FT TRANSMEM 8241 8260
 FT TRANSMEM 8261 8280
 FT TRANSMEM 8281 8300
 FT TRANSMEM 8301 8320
 FT TRANSMEM 8321 8340
 FT TRANSMEM 8341 8360
 FT TRANSMEM 8361 8380
 FT TRANSMEM 8381 8400
 FT TRANSMEM 8401 8420
 FT TRANSMEM 8421 8440
 FT TRANSMEM 8441 8460
 FT TRANSMEM 8461 8480
 FT TRANSMEM 8481 8500
 FT TRANSMEM 8501 8520
 FT TRANSMEM 8521 8540
 FT TRANSMEM 8541 8560
 FT TRANSMEM 8561 8580
 FT TRANSMEM 8581 8600
 FT TRANSMEM 8601 8620
 FT TRANSMEM 8621 8640
 FT TRANSMEM 8641 8660
 FT TRANSMEM 8661 8680
 FT TRANSMEM 8681 8700
 FT TRANSMEM 8701 8720
 FT TRANSMEM 8721 8740
 FT TRANSMEM 8741 8760
 FT TRANSMEM 8761 8780
 FT TRANSMEM 8781 8800
 FT TRANSMEM 8801 8820
 FT TRANSMEM 8821 8840
 FT TRANSMEM 8841 8860
 FT TRANSMEM 8861 8880
 FT TRANSMEM 8881 8900
 FT TRANSMEM 8901 8920
 FT TRANSMEM 8921 8940
 FT TRANSMEM 8941 8960
 FT TRANSMEM 8961 8980
 FT TRANSMEM 8981 9000
 FT TRANSMEM 9001 9020
 FT TRANSMEM 9021 9040
 FT TRANSMEM 9041 9060
 FT TRANSMEM 9061 9080
 FT TRANSMEM 9081 9100
 FT TRANSMEM 9101 9120
 FT TRANSMEM 9121 9140
 FT TRANSMEM 9141 9160
 FT TRANSMEM 9161 9180
 FT TRANSMEM 9181 9200
 FT TRANSMEM 9201 9220
 FT TRANSMEM 9221 9240
 FT TRANSMEM 9241 9260
 FT TRANSMEM 9261 9280
 FT TRANSMEM 9281 9300
 FT TRANSMEM 9301 9320
 FT TRANSMEM 9321 9340
 FT TRANSMEM 9341 9360
 FT TRANSMEM 9361 9380
 FT TRANSMEM 9381 9400
 FT TRANSMEM 9401 9420
 FT TRANSMEM 9421 9440
 FT TRANSMEM 9441 9460
 FT TRANSMEM 9461 9480
 FT TRANSMEM 9481 9500
 FT TRANSMEM 9501 9520
 FT TRANSMEM 9521 9540
 FT TRANSMEM 9541 9560
 FT TRANSMEM 9561 9580
 FT TRANSMEM 9581 9600
 FT TRANSMEM 9601 9620
 FT TRANSMEM 9621 9640
 FT TRANSMEM 9641 9660
 FT TRANSMEM 9661 9680
 FT TRANSMEM 9681 9700
 FT TRANSMEM 9701 9720
 FT TRANSMEM 9721 9740
 FT TRANSMEM 9741 9760
 FT TRANSMEM 9761 9780
 FT TRANSMEM 9781 9800
 FT TRANSMEM 9801 9820
 FT TRANSMEM 9821 9840
 FT TRANSMEM 9841 9860
 FT TRANSMEM 9861 9880
 FT TRANSMEM 9881 9900
 FT TRANSMEM 9901 9920
 FT TRANSMEM 9921 9940
 FT TRANSMEM 9941 9960
 FT TRANSMEM 9961 9980
 FT TRANSMEM 9981 10000
 FT TRANSMEM 10001 10020
 FT TRANSMEM 10021 10040
 FT TRANSMEM 10041 10060
 FT TRANSMEM 10061 10080
 FT TRANSMEM 10081 10100
 FT TRANSMEM 10101 10120
 FT TRANSMEM 10121 10140
 FT TRANSMEM 10141 10160
 FT TRANSMEM 10161 10180
 FT TRANSMEM 10181 10200
 FT TRANSMEM 10201 10220
 FT TRANSMEM 10221 10240
 FT TRANSMEM 10241 10260
 FT TRANSMEM 10261 10280
 FT TRANSMEM 10281 10300
 FT TRANSMEM 10301 10320
 FT TRANSMEM 10321 10340
 FT TRANSMEM 10341 10360
 FT TRANSMEM 10361 10380
 FT TRANSMEM 10381 10400
 FT TRANSMEM 10401 10420
 FT TRANSMEM 10421 10440
 FT TRANSMEM 10441 10460
 FT TRANSMEM 10461 10480
 FT TRANSMEM 10481 10500
 FT TRANSMEM 10501 10520
 FT TRANSMEM 10521 10540
 FT TRANSMEM 10541 10560
 FT TRANSMEM 10561 10580
 FT TRANSMEM 10581 10600
 FT TRANSMEM 10601 10620
 FT TRANSMEM 10621 10640
 FT TRANSMEM 10641 10660
 FT TRANSMEM 10661 10680
 FT TRANSMEM 10681 10700
 FT TRANSMEM 10701 10720
 FT TRANSMEM 10721 10740
 FT TRANSMEM 10741 10760
 FT TRANSMEM 10761 10780
 FT TRANSMEM 10781 10800
 FT TRANSMEM 10801 10820
 FT TRANSMEM 10821 10840
 FT TRANSMEM 10841 10860
 FT TRANSMEM 10861 10880
 FT TRANSMEM 10881 10900
 FT TRANSMEM 10901 10920
 FT TRANSMEM 10921 10940
 FT TRANSMEM 10941 10960
 FT TRANSMEM 10961 10980
 FT TRANSMEM 10981 11000
 FT TRANSMEM 11001 11020
 FT TRANSMEM 11021 11040
 FT TRANSMEM 11041 11060
 FT TRANSMEM 11061 11080
 FT TRANSMEM 11081 11100
 FT TRANSMEM 11101 11120
 FT TRANSMEM 11121 11140
 FT TRANSMEM 11141 11160
 FT TRANSMEM 11161 11180
 FT TRANSMEM 11181 11200
 FT TRANSMEM 11201 11220
 FT TRANSMEM 11221 11240
 FT TRANSMEM 11241 11260
 FT TRANSMEM 11261 11280
 FT TRANSMEM 11281 11300
 FT TRANSMEM 11301 11320
 FT TRANSMEM 11321 11340
 FT TRANSMEM 11341 11360
 FT TRANSMEM 11361 11380
 FT TRANSMEM 11381 11400
 FT TRANSMEM 11401 11420
 FT TRANSMEM 11421 11440
 FT TRANSMEM 11441 11460
 FT TRANSMEM 11461 11480
 FT TRANSMEM 11481 11500
 FT TRANSMEM 11501 11520
 FT TRANSMEM 11521 11540
 FT TRANSMEM 11541 11560
 FT TRANSMEM 11561 11580
 FT TRANSMEM 11581 11600
 FT TRANSMEM 11601 11620
 FT TRANSMEM 11621 11640
 FT TRANSMEM 11641 11660
 FT TRANSMEM 11661 11680
 FT TRANSMEM 11681 11700
 FT TRANSMEM 11701 11720
 FT TRANSMEM 11721 11740
 FT TRANSMEM 11741 11760
 FT TRANSMEM 11761 11780
 FT TRANSMEM 11781 11800
 FT TRANSMEM 11801 11820
 FT TRANSMEM 11821 11840
 FT TRANSMEM 11841 11860
 FT TRANSMEM 11861 11880
 FT TRANSMEM 11881 11900
 FT TRANSMEM 11901 11920
 FT TRANSMEM 11921 11940
 FT TRANSMEM 11941 11960
 FT TRANSMEM 11961 11980
 FT TRANSMEM 11981 12000
 FT TRANSMEM 12001 12020
 FT TRANSMEM 12021 12040
 FT TRANSMEM 12041 12060
 FT TRANSMEM 12061 12080
 FT TRANSMEM 12081 12100
 FT TRANSMEM 12101 12120
 FT TRANSMEM 12121 12140
 FT TRANSMEM 12141 12160
 FT TRANSMEM 12161 12180
 FT TRANSMEM 12181 12200
 FT TRANSMEM 12201 12220
 FT TRANSMEM 12221 12240
 FT TRANSMEM 12241 12260
 FT TRANSMEM 12261 12280
 FT TRANSMEM 12281 12300
 FT TRANSMEM 12301 12320
 FT TRANSMEM 12321 12340
 FT TRANSMEM 12341 12360
 FT TRANSMEM 12361 12380
 FT TRANSMEM 12381 12400
 FT TRANSMEM 12401 12420
 FT TRANSMEM 12421 12440
 FT TRANSMEM 12441 12460
 FT TRANSMEM 12461 12480
 FT TRANSMEM 12481 12500
 FT TRANSMEM 12501 12520
 FT TRANSMEM 12521 12540
 FT TRANSMEM 12541 12560
 FT TRANSMEM 12561 12580
 FT TRANSMEM 12581 12600
 FT TRANSMEM 12601 12620
 FT TRANSMEM 12621 12640
 FT TRANSMEM 12641 12660
 FT TRANSMEM 12661 12680
 FT TRANSMEM 12681 12700
 FT TRANSMEM 12701 12720
 FT TRANSMEM 12721 12740
 FT TRANSMEM 12741 12760
 FT TRANSMEM 12761 12780
 FT TRANSMEM 12781 12800
 FT TRANSMEM 12801 12820
 FT TRANSMEM 12821 12840
 FT TRANSMEM 12841 12860
 FT TRANSMEM 12861 12880
 FT TRANSMEM 12881 12900
 FT TRANSMEM 12901 12920
 FT TRANSMEM 12921 12940
 FT TRANSMEM 12941 12960

FT BINDING 307 307 HEME 4 (COVALENT) (BY SIMILARITY).
 FT BINDING 310 310 HEME 4 (COVALENT) (BY SIMILARITY).
 FT METAL 311 311 IRON 4 (HEME AXIAL LIGAND)
 FT METAL 311 311 (BY SIMILARITY).
 SQ SEQUENCE 383 AA; 41522 MW; 96BCD91EF1B9AE7E CRC64;
 Query Match 39.4%; Score 43; DB 1; Length 383;
 Best Local Similarity: 36.8%; Pred. No. 8.9;
 Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
 QY 1 TPDINPAMXXRGIRPYGR 19
 DB 263 TPORTTAMVARIHVVDINQ 281
 YAAJ ECOLI STANDARD; PRT; 476 AA.
 ID YAAJ ECOLI STANDARD; PRT; 476 AA.
 AC P30143;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative transporter yaaJ.
 GN YAAJ OR B0007.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 ON NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=92334977; PubMed=1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
 RA Isono K., Mizobuchi K., Nakata A.,
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 0-2.4 min region."
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SODIUM/ALANINE SYMPORTER FAMILY
 (SAF). STRONG, TO H. INFLUENZAE H10183.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: D10483; -; NOT ANNOTATED_CDS.
 DR EMBL: AE000111; AAC73118.1; -;
 DR Ecogene: EGI1555; yaaJ.
 DR InterPro: IPR002293; AA/rel_prmaseel.
 DR InterPro: IPR001463; Na_Ala_symptr.
 DR Pfam: PF01235; Na_Ala_symp.1.
 DR PRINTS: PR00175; NAALASMPOR.
 DR TIGRfam: TIGR00835; agcs.1.
 DR PROSITE: PS00873; NA_ALANINE_SYM.1.
 KW Hypothetical protein; Transmembrane; Inner membrane; Transport;
 FT Symport; Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 81 101 POTENTIAL.

FT TRANSMEM 141 161 POTENTIAL.
 FT TRANSMEM 174 194 POTENTIAL.
 FT TRANSMEM 207 227 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 300 320 POTENTIAL.
 FT TRANSMEM 351 371 POTENTIAL.
 FT TRANSMEM 391 411 POTENTIAL.
 FT TRANSMEM 414 434 POTENTIAL.
 SQ SEQUENCE 476 AA; 51662 MW; 2F6B2E12E126E63 CRC64;
 Query Match 39.4%; Score 43; DB 1; Length 476;
 Best Local Similarity: 44.4%; Pred. No. 11;
 Matches 8; Conservative 2; Mismatches 2; Indels 6; Gaps 1;
 QY 3 DIN-----PAMYXXRGI 14
 DB 120 DVNGQFRGGPAMTAMRGL 137
 YAAJ YEAST STANDARD; PRT; 719 AA.
 ID NRP1 YEAST STANDARD; PRT; 719 AA.
 AC P32770; Q12228;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Asparagine-rich protein (ARP protein).
 GN NRP1 OR ARP1 OR ARP OR YDL167C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AH22;
 RX MEDLINE=93247548; PubMed=8483449;
 RA Wehner E.P., Rao E., Brendel M.,
 RT "Molecular structure and genetic regulation of SPA, a gene
 responsible for resistance to formaldehyde in Saccharomyces
 cerevisiae, and characterization of its protein product."
 RL Mol. Gen. Genet. 237:351-358(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288c;
 RX Pohl T.M.,
 RT Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: X68020; CAA48159.1; -;
 DR EMBL: Z67750; CAA91579.1; -;
 DR EMBL: Z74215; CAA98741.1; -;
 DR PIR: S31139; S31139.
 DR SGI: S0002326; NRP1.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR InterPro: IPR001876; Znf_RangBP.
 DR Pfam: PF00076; rrm.1.
 DR Pfam: PF00641; zf_RanBP.2.
 DR SMART: SM00360; RRM.1.
 DR SMART: SM00347; Znf_RBZ.2.
 DR PROSITE: PS50102; RRM.1.
 DR PROSITE: PS00030; RRM_RNP.1; FALSE_NEG.
 DR PROSITE: PS01358; ZF_RANBP2.1; 2.
 DR PROSITE: PS50199; ZF_RANBP2.2; 2.
 KW Nuclear protein; Zinc-finger; RNA-binding; Repeat.

Query Match
Best Local Similarity 39.4%; Score 43; DB 1; Length 719;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DOMAIN 226 322 RNA-BINDING (RRM).
ZNF226 355 384 RANBP2-TYPE 1.
ZNF226 581 610 RANBP2-TYPE 2.
DOMAIN 490 564 ASH-RICH.
CONFLICT 493 493 I -> N (IN REF. 1).
SEQUENCE 719 AA; 79299 MW; ADA9BC09FD582669 CRC64;

Query Match
Best Local Similarity 50.0%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

7 AMYXXRGIRPVG 18
: : : : :
244 SMFTQYGVRPV 255

RESULT 10
EX7L_STRCO STANDARD; PRT; 402 AA.
ID EX7L_STRCO
AC 09EBM3;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR SC05056 OR SCK7.29C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; Pubmed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RL Nature 417:141-147(2002).
CC -1- FUNCTION: BIDIRECTIONALLY DEGRADABLE SINGLE-STRANDED DNA INTO LARGE
ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5' to 3' or
3' to 5'-direction to yield nucleoside 5'-phosphates.
CC -1- SUBUNIT: HETEROOLIGOMER COMPOSED OF LARGE AND SMALL SUBUNITS (BY
SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: AL391754; CAC05901.1;
DR InterPro: IPR003753; Exonuc_VII_L.
DR InterPro: IPR004365; RNA_ant1.
DR Pfam: PF01336; RNA_ant1; 1.
DR Pfam: PF02601; Exonuc_VII_L; 1.
DR TIGRFAMs: TIGR00237; xsea; 1.
KW Hydrolyase; Nuclease; Exonuclease; Complete proteome.
DE SEQUENCE 402 AA; 43882 MW; 145929A8372B4E08 CRC64;

Query Match
Best Local Similarity 38.5%; Score 42; DB 1; Length 402;
Matches 9; Conservative 1; Mismatches 3; Indels 8; Gaps 1;

6 PAMYXXRG-----IRPV 18
: : : : :
89 PEMYAPRGCLSLRAAEIRPV 109

RESULT 11
ARGC_BACHD STANDARD; PRT; 345 AA.
ID ARGC_BACHD
AC 09X8V2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-
acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).
GN ARGC OR BH2900.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; Pubmed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Holikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+) +
phosphate = N-acetyl-5-glutamyl phosphate + NADPH.
CC -1- PATHWAY: Arginine biosynthesis; third step.
CC -1- SIMILARITY: BELONGS TO THE NAGSA DEHYDROGENASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: AP001517; BAB06619.1;
DR InterPro: IPR000706; AGPR_act_site.
DR InterPro: IPR000534; Semialdh_dh.
DR Pfam: PF01118; Semialdh_dh; 1.
DR Pfam: PF02774; Semialdh_dh; 1.
DR ProDom: PD003765; AGPR_act_site; 1.
DR PROSITE: PS01224; ARGC; 1.
KW Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.
FT ACT_SITE 149 149 BY SIMILARITY
SEQUENCE 345 AA; 38188 MW; 3E9F45DD09FC08EA CRC64;

Query Match
Best Local Similarity 38.1%; Score 41.5; DB 1; Length 345;
Matches 9; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

2 PDINPAMYXXRGIRPV 18
: : : : :
106 PDYEAMV-KROAAPV 121

RESULT 12
Y762_METJA STANDARD; PRT; 342 AA.
ID Y762_METJA
AC 058172;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M00762.

```

GN      MJ0762      Methanococcus jannaschii.
OS
OC      Archaeae: Euryarchaeota, Methanococci; Methanococcales;
OC      Methanocaldococcales; Methanocaldococcus.
OX      NCBI_TaxID=2190;
RN
RM      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX      MEDLINE=96337999; PubMed=8668087;
RA      Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA      Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams C.I.,
RA      Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA      Scott J.L., Geoghegan N.S.M., Weisman J.F., Fuhlmann J.L., Nguyen D.,
RA      Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA      Klenk H.-P., Fraser C.M., Smith H.O., Weese C.R., Venter J.C.;
RT      "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT      jannaschii."
RL      Science 273:1058-1073(1996).
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential)
CC      -1- SIMILARITY: STRONG, TO M.JANNASCHII MJ0576 AND TO S.POMBE MALARE
CC      PENSEME (MAL1).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U67521; AAB98753.1; -.
DR      TIGR; MJ0762; -.
DR      InterPro; IPR004695; C4dic_mal_transp.
DR      Pfam; PF03595; C4dic_mal_tran; 1.
DR      TIGRFAMs; TIGR00816; tdt; 1.
KW      Hypothetical protein; Transmembrane; Transport; Complete proteome.
FT      TRANSMEM 8
FT      TRANSMEM 28
FT      TRANSMEM 39
FT      TRANSMEM 59
FT      TRANSMEM 78
FT      TRANSMEM 99
FT      TRANSMEM 108
FT      TRANSMEM 128
FT      TRANSMEM 142
FT      TRANSMEM 162
FT      TRANSMEM 175
FT      TRANSMEM 195
FT      TRANSMEM 207
FT      TRANSMEM 227
FT      TRANSMEM 242
FT      TRANSMEM 262
FT      TRANSMEM 276
FT      TRANSMEM 296
FT      TRANSMEM 304
FT      TRANSMEM 324
SQ      SEQUENCE 342 AA; 395534 MW; 08EFEC3E2C4955D8 CRC64;
Query Match 37.6%; Score 41; DB 1; Length 342;
Best Local Similarity 53.3%; Pred. No. 17;
Matches 8; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
QY 4 INPAXXXRGIRPPV 18
Db 139 VNPGRY-----IPVG 149

RESULT 13
ID Y576.METJA STANDARD; PRT; 347 AA.
AC 057996;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0576.
GN MJ0576.
OS Methanococcus jannaschii.
OC Archaeae: Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
NN [1]

```

```

RP SEQUENCEFROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Meldrum J.F., Fuhrman J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith R.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: STRONG, TO M.JANNSCHIT MJ0762 AND TO S.POMBE MALATE
CC PERMEASE (MAE1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, U67506; AAB98567.1; -.
DR TIGR; MJ0576; -.
DR InterPro: IPR004695; Cddic_mal_transp.
DR Pfam: PF03595; Cddic_mal_tran; 1.
DR TIGRFAMS; TIGR00816; tdt; 1.
DR Hypothetical protein; Transmembrane; Transport; Complete proteome.
KT TRANSMEM 15 35
FT TRANSMEM 46 66 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 249 269 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
SQ SEQUENCE 347 AA; 39556 MW; 632F671A31DE183 CMC64;

Query Match 37.68; Score 41; DB 1; Length 347;
Best Local Similarity 53.3%; Pred. No. 17;
Matches 8; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 4 INPAMYXXRGIRPVG 18
Db 146 VNPGRY-----IPVG 156

RESULT 14
CYGR_ARBP CYGR_ARBP STANDARD; PRT; 986 AA.
AC P11528;
DF 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE Resact receptor precursor (Guanylate cyclase) (EC 4.6.1.2).
OS Arabacia punctulata (Punctate sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinioidea; Euechinoidea; Echinacea; Arbacoidea; Arbaciidae; Arbacia.
OX NCBI_TaxID=7641;
RN [1]
RT SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Testis;
RA MEDLINE=88318927; PubMed=2901039;
RA Slingh S., Lowe D.G., Thorpe D.S., Rodriguez H., Kuang W.-J.,
RA Dangott L.J., Chinkens M., Goeddel D.V., Garbers D.L.;
RT "Membrane guanylate cyclase is a cell-surface receptor with homology
to protein kinases."

```

RL Nature 334:708-712(1988).
 CC -1- FUNCTION: IMPLICATED AS A CELL-SURFACE RECEPTOR ON SPERMATOZOIA
 CC FOR 'RESCAT' A CHEMOTACTIC PEPTIDE, AND ON VARIOUS OTHER CELLS
 CC AS A RECEPTOR FOR ATRIAL NATRIURETIC PEPTIDE.
 CC -1- CATALYTIC ACTIVITY: GTP -> 3',5'-cyclic GMP + diphosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X12874; CNA31367.1; -.
 DR PIR: S05480; OYURCA.
 DR InterPro: IPR001828; ANF_receptor.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001054; G_cyclase.
 DR Pfam: PF000659; kinase_1.
 DR Pfam: PF00211; guanylate_cyc: 1.
 DR Pfam: PF01094; ANF_receptor; 1.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR ProSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Receptor; Transmembrane; Glycoprotein; Phosphorylation; Lyase;
 KW cAMP synthesis; Signal.
 KM SIGNAL 1 21
 FT CHAIN 1 22 986
 FT DOMAIN 22 507 RESACT RECEPTOR.
 FT TRANSMEM 508 528 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 529 986 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 568 836 PROTEIN KINASE LIKE.
 FT CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 986 AA; 111284 MW; B40238A74CCARC52 CRC64;
 Query Match 37.6%; Score 41; DB 1; Length 986;
 Best Local Similarity 62.5%; Pred. No. 47;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PDINPAMY 9
 DB 475 PDLNPMVH 482
 RESULT 15
 Y493_MYCTU STANDARD; PRT; 329 AA.
 ID Y493_MYCTU
 AC 01158;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV0493c.
 GN RV0493c OR MT0513 OR MTCY2069.19C.
 OS Mycobacterium tuberculosis.
 CC Bacteria: Actinobacteria: Actinobacteria (class): Actinobacteridae;
 CC Actinomycetales: Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CC NCBI_TaxID=1773;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=H37Rv;
 RC MEDLINE=98295987; PubMed=9634230;
 RX Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagsels K., Kirogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikala A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: SOME, TO MLEPRAE B2168.F2.93.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z77162; CAB00954.1; -.
 DR EMBL: AE006952; AAK44736.1; -.
 DR TIGR: MT0513; -.
 DR Tuberculist; RV0493c; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 329 AA; 35427 MW; BE4B524750277B56 CRC64;
 Query Match 36.7%; Score 40; DB 1; Length 329;
 Best Local Similarity 42.9%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 QY 2 PDINPAMYXXRCIR 15
 DB 90 PAAGPAMFDIACVR 103

Search completed: May 1, 2003, 14:33:20
 Job time : 8.21875 secs

Thu May 1 15:54:11 2003

seq74-plus-73.rsp

SEQ ID NO: 74 fused to SEQ ID NO: 73
Database: SwissProt-40
AC NO: P81278

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:36:27 ; Search time 11 Seconds
(without alignments)

124.429 Million cell updates/sec

Title: SEQ74-PLUS-73

Perfect score: 136
Sequence: 1 SRXHXSMKXKRPDPINPAMYXXNGIRPVGRFX 33

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	147	94.2	83	PRRP_RAT	P81278 rattus norv
2	147	94.2	87	PRRP_HUMAN	P81277 homo sapien
3	146	93.6	98	PRRP_BOVIN	P81264 bos taurus
4	65	41.7	4486	DYH9_HUMAN	O979c3 homo sapien
5	62	39.7	2476	ZAN_PIG	Q28983 sus scrofa
6	62	39.7	2832	NDVB_RHINE	P20471 rhizobium m
7	61	39.1	999	HGP4_HAEIN	O57408 haemophilus
8	61	39.1	1997	OTOF_HUMAN	O99c10 homo sapien
9	61	39.1	1997	OTOF_MOUSE	O99c11 mus musculu
10	60	38.5	873	FAS2_DROME	P34082 drosophila
11	60	38.5	962	UVRA_METH	O26543 methanobact
12	60	38.5	3866	HRX_MOUSE	P05200 mus musculu
13	60	38.5	3969	HRX_HUMAN	O03164 homo sapien
14	60	38.5	4393	PCBM_HUMAN	P98160 homo sapien
15	60	38.5	4568	DYH8_CHIRE	O38565 chlamydomon
16	59	37.8	371	GALS_MOUSE	O88854 mus musculu
17	59	37.8	372	GALS_RAT	O08726 rattus norv
18	59	37.8	637	MTHA_DROME	O940r5 drosophila
19	59	37.8	940	UVRA_ECOLI	O858u9 escherichia
20	59	37.8	940	UVRA_ECOLI	O07671 escherichia
21	59	37.8	941	UVRA_SALTY	P37434 salmonella
22	59	37.8	1138	C7AB_BACTO	O03749 bacillus th
23	59	37.8	1138	C7AB_BACTO	O45707 bacillus th
24	59	37.8	1555	GDE_RABIT	P33574 o glycoen
25	59	37.8	3341	PGLG_MCF4	P23098 genome po
26	59	37.8	4466	DYH8_TRIGR	O33040 b tyrocidin
27	59	37.8	6486	TYOC_BACBR	P12894 mouse intra
28	58	37.2	814	POL_IPMAI	O9kuw5 vibrio chol
29	58	37.2	960	UVRA_VIBCH	O83527 treponema p
30	58	37.2	940	UVRA_TREPA	P34731 c fatty aci
31	58	37.2	2225	FAST_CANAL	P08955 mesocricetu
32	58	37.2	2225	PVRL_MESAU	P98081 drosophila
33	58	37.2	2411	DAB_DROME	

34	58	37.2	2715	1	TRX2_HUMAN	Q9um66 homo sapien
35	58	37.2	4590	1	FATH_HUMAN	Q14517 homo sapien
36	58	37.2	4590	1	GALS_HUMAN	O43603 homo sapien
37	57	36.5	491	1	ADRO_HUMAN	P22570 homo sapien
38	57	36.5	867	1	POL_IPMA	P13368 mouse intra
39	57	36.5	905	1	NIA_FUSOX	P39863 fusarium ox
40	57	36.5	926	1	UVRA_AQUAE	O66911 aquifex aeo
41	57	36.5	986	1	CYGR_ARBPV	P1528 arabacia pun
42	57	36.5	1073	1	YFPA_YEAST	P43564 saccharomyc
43	57	36.5	1165	1	POL_GALV	P21414 gibbon ape
44	57	36.5	1532	1	GDE_HUMAN	P35573 h glycogen
45	57	36.5	1636	1	PKR5_BRACL	Q9nj15 branchiosto

ALIGNMENTS

RESULT 1	PRRP_RAT	STANDARD	PRT	83 AA.
AC	P81278			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Prolactin-releasing peptide precursor (PrRP) (Prolactin-releasing hormone) [Contains: Prolactin-releasing peptide PrRP31; Prolactin-releasing peptide PrRP20].			
GN	PrRP.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RC	MEDLINE=98268781; PubMed=9607765;			
RA	Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kikada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.; "A prolactin-releasing peptide in the brain.";			
RT	Nature 393:272-276(1998).			
RL	[2]			
RP	TISSUE SPECIFICITY. Pubmed=10498338; MEDLINE=9942652;			
RA	Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kikada C., Kurokawa T., Nishimura O., Onda H., Sunino Y., Fujino M.; "Tissue distribution of prolactin-releasing peptide (PrRP) and its receptor.";			
RT	Regul. Pept. 83:1-10(1999).			
RL	Regul. Pept. 83:1-10(1999).			
CC	-1- FUNCTION: Stimulates prolactin (PR) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PrL.			
CC	-1- TISSUE SPECIFICITY: Widely expressed, with highest levels in medulla oblongata and hypothalamus.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AB015418; BAA29026.1;			
KW	Hormone; Amidation; Signal; Cleavage on pair of basic residues.			
FT	SIGNAL	21		
FT	PEPTIDE	22		
FT	PEPTIDE	33		
FT	MOD RES	52		
FT	SEQUENCE	83 AA; 9215 MW; DOC75A264BEE4F29 CRC64;		
Query Match	94.2%	Score 147;	DB 1;	Length 83;

RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Nasal epithelium;
 RA Bartoloni L., Blouin J.L., Maiti A., Sainsbury A., Rossier C.,
 RA Gehrig C., She J.X., Marron M.P., Lander M.P., Meeks M., Chung E.,
 RA Jorissen M., Scott H.S., Delozier Blanchet C.D., Gardiner M.,
 RA Antonarakis S.E.;
 RT "Axonemal beta heavy chain dynein DNAH9: cDNA sequence, genomic
 RT structure and investigation of its role in primary ciliary
 RT dyskinesia.";
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1874-1974 FROM N.A.
 RC TISSUE-Nasal polyps;
 RA Maiti A.K., Maiti M.G., Jorissen M., Volz A., Ziegler A.,
 RA Bouvagnet P.;
 RT "Chromosomal localization of human dynein heavy chain genes.";
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=20558134; PubMed=11104725;
 RA Reed W., Carson J.L., Moats-Staats B.M., Lucier T., Hu P.C.,
 RA Brighton L., Gambing T.M., Huang C.H., Leigh M.W., Collier A.M.;
 RT "Characterization of an axonemal dynein heavy chain expressed early in
 RT airway epithelial ciliogenesis.";
 RL Am. J. Respir. Cell Mol. Biol. 23:734-741(2000).
 CC -1- FUNCTION: FORCE GENERATING PROTEIN OF RESPIRATORY CILIA. PRODUCES
 CC FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES. DYNEIN HAS ATPASE
 CC ACTIVITY.
 CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
 CC INTERMEDIATE AND LIGHT CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE257737; AAF69004.1; -
 CC EMBL: AJ404468; CAB94756.1; -
 CC EMBL: AJ133088; CA10561.1; -
 CC Gene: HGNC:2953; DNAH9.
 DR MIM: 603330; -
 DR InterPro: IPR004273; Dynein_heavy.
 DR Pfam: PF03028; Dynein_heavy; 1.
 KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
 FT DOMAIN 381 410 COILED COIL (POTENTIAL).
 FT DOMAIN 504 529 COILED COIL (POTENTIAL).
 FT DOMAIN 639 662 COILED COIL (POTENTIAL).
 FT DOMAIN 752 823 COILED COIL (POTENTIAL).
 FT DOMAIN 1326 1355 COILED COIL (POTENTIAL).
 FT DOMAIN 3051 3154 COILED COIL (POTENTIAL).
 FT DOMAIN 3285 3341 COILED COIL (POTENTIAL).
 FT DOMAIN 3640 3675 COILED COIL (POTENTIAL).
 FT NP_BIND 1870 1877 ATP (POTENTIAL).
 FT NP_BIND 2151 2158 ATP (POTENTIAL).
 FT NP_BIND 2478 2485 ATP (POTENTIAL).
 FT NP_BIND 2825 2832 ATP (POTENTIAL).
 FT CONFLICT 2505 2505 L -> V (IN REF. 2).
 FT CONFLICT 3678 3678 T -> A (IN REF. 2).
 FT CONFLICT 4374 4374 I -> M (IN REF. 2).
 SQ SEQUENCE 4486 AA; 511927 MW; 996BDFDEB0B3EB1 CRC64;

Query Match 41.7%; Score 65; DB 1; Length 4486;
 Best Local Similarity 0.6%; Pred. No. 5.4e+02;
 Matches 19; Conservative 1; Mismatches 10; Indels 3343; Gaps 5;
 2 RXHXSM----- 8

Db 163 RRAHSLQCDLSVILEQVKGKTLPLPAGSEKMEPADSKSETVIDSDKSVIAIESAVI 222
 QY 9 ----- 8
 Db 223 KMSYQVQVVKRESSQPLLOGENPTPKVELEFWMSRYEDLKYYINQRLRTTVRGMAKILD 282
 QY 9 ----- 8
 Db 283 KLOSSFPAFAMRYDVAALAEADIHVHLIPQRHLEALENAEPFVKPOLRPLHHV 342
 QY 9 ----- 8
 Db 343 CLIMATCKSYSPGRILVLOEICNLILQASNTLSPEDLIRSEVSQRRLQVSDPTLS 402
 QY 9 ----- 8
 Db 403 PFKEQFQDRRENLTFFKENEQVKEWDFQSSLVYRLDGLGRLLHYVGLTKALDTHKL 462
 QY 9 ----- 8
 Db 463 GKVEFSGVRGNALSOQVOOMEHEFOEMTRLLSGSSDCLYQSTDFENDVSEFNQVEDL 522
 QY 9 ----- 8
 Db 523 DRRLLTITLQAFDAPGLEHAFKLIDLAGNLLERPLVARDTSKYLVLIQMFNKDLAYR 582
 QY 9 ----- 8
 Db 583 MYSQVQEAELGSPYHKMNPVAGGLRMAQLRQIRQPSFNFRITHPCMSAEGK 642
 QY 9 ----- 8
 Db 643 RMQKREDMLSLEKETRYLWEDMCRTVSEKSYNLSQPLKRPETKEITINENPOLIS 702
 QY 9 ----- 8
 Db 703 VLKEMSYLEPRKMHMPETAAMSSRDFTQYLVANLEMANMTNKYKTLLEVEPLVE 762
 QY 9 ----- 8
 Db 763 EELQNLRLRAEETLNMKTEGICDYVEITSSIHLEQRIOKTKDNVEIOMKMTW 822
 QY 9 ----- 8
 Db 823 TPFRKDKKRESLSLDHRDMKEYNLLKESGLKTHALVOENGLFSADPTSNIMKT 882
 QY 9 ----- 8
 Db 883 YVNSIDNLLNGFLAIECSLKYLLNTECKAGLTPFPAQLSLAIPELVYPSLESQVK 942
 QY 9 ----- 8
 Db 943 GGFCDIVELGITSIFRIPSLVPLSPQNSPHYQVDLGIPLDANMRTLMERYQRMGL 1002
 QY 9 ----- 8
 Db 1003 CCGVQSTFSQSYLYVEDRKEVLAQFLLYGHILPPEIEDHVEDGIEBNPLLSQFRVQI 1062
 QY 9 ----- 8
 Db 1063 DSYETLYEVCRLPEIKVFDGMMKIDIRPFKASLLNIKKRSLFLKQHLVDHTHSIANTL 1122
 QY 9 ----- 8
 Db 1123 DAFIKSESGLLKVEKDFQGLVEIMGHMAVEROSNTDEMEPEPKOTIELTKTYEQE 1182
 QY 9 ----- 8
 Db 1183 LPEYVFKOLELPEKMNIRKVAITVAKQVAPLOANEVTLIRORCTAPDAEQQFWBOFH 1242
 QY 9 ----- 14
 Db 1243 KEAPRFDSIPHQMLDARHIEIQOMSTMASISESASLFEVNVVDYKQLRQCKREVCQL 1302

```

QY 15 ----- 14
Db 1303 KELMDTIGMTSSIHANETTPMNRINVEAMELECKOFARHIRNLDKEVRAMDAFTGIEST 1362
QY 15 ----- 14
Db 1363 VMNTLSLRAVAELQNPABIRERHWRQMQATGVSTMDQDTLAHLQLOLHHYEDEVRG 1422
QY 15 ----- 14
Db 1423 IVDKAKEMGKTKLKELOTTWAGMEFOYERHPRTNPLCSDEDLIEVLENOVOLQNL 1482
QY 15 ----- 14
Db 1483 VMSKYVAFLEEVSGMOKLSTVDVAVISIMFEVQRTWHLESIFTSSEDIRAOLPODSKR 1542
QY 15 ----- 14
Db 1543 FEGIDIDFKELAYDAQKIPNVVOTNKGLEYEKLEDIQGRCLCEKALAEYLDTKRLAEP 1602
QY 15 ----- 14
Db 1603 RFTYLSDDLILSNGTAPQOVRHLSKLFDNNAKRFQDLSGSEPTSLGNTSKKEE 1662
QY 15 ----- 14
Db 1663 YVAFSEPODSGOVEIMLNHVLGHMKATVHEMTEGVTAYEERPREQWLFDRQAVALTC 1722
QY 15 ----- 14
Db 1723 TQIMWTEVGMAFARLEEGESAMKDYKKOVALKTLITMLIGOLSGDRORIMTCTI 1782
QY 15 ----- 14
Db 1783 DVHARDVYAKMIQKVNAQAFMLSQLRHRWDEYKHCENICDAQFLXYEYLGNTPR 1842
QY 15 ----- 14
Db 1843 LVTPPLDRCYITLQSLHLTMGAPAGAPAGTGTETTKDGRALGILVYVNCSEQMDY 1902
QY 15 ----- 14
Db 1903 KSCGNITKGLAQMGCFDEFENRISVEVLSVAVQYSIODAIRDKKOWSEFLGEEISL 1962
QY 15 ----- 14
Db 1963 NPSVGITTMNPGYAGTELPENLKSIFRCAMVVPDEFELICEIMLVAEGFIEAQSLAK 2022
QY 15 ----- 14
Db 2023 FITLYOLCKELLSKODHYDWGLRAIKSVLVAGSLKRGDPREDQVLMRLDNFNPKI 2082
QY 15 ----- 14
Db 2083 VTDDMPILFMGLIGLFPALDVPRRRDNPEFALYRKALVDLKLQADNFVLKVVQLEELLA 2142
QY 15 ----- 14
Db 2143 VRHSVFPVVGAGTGSQVLSLHKTQIMKRPRVTDLNRKATYNDELFGIINPATGEMK 2202
QY 15 ----- 14
Db 2203 DGLFSSIMRELANTHDGPKWILLDGDIDPMWIESLNTVMDNKNVLTLASNERILPLNPM 2262
QY 15 ----- 18
Db 2263 KLFELISLRTAPATVSRAGIILYINPADLGWNPVSSWIEKREIQTERANLTLEDKYL 2322
QY 19 ----- 18
Db 2323 PTCIDTLRTFRPKTIIPBESQMVOMVCHLLECLITTEDIPADCPKEIYEYFVFAIMAR 2382

```

```

QY 19 ----- 18
Db 2383 GGAMVODOLVDYRAEFSSKMWLEFETVAFPSOGTTFEDYIDPETKKFERPSKLVPOFED 2442
QY 19 ----- 18
Db 2443 PEMPLQACLVHTSETIRVCYEMERLARQRPVMLVGTAGTGSVLGAKLASLDEAVLY 2502
QY 19 ----- 18
Db 2503 KULPNTYTTSSMLQAVLEKPLEKAGRNYPGNKKLIYFIDMNMPEVDAVGTVOEHT 2562
QY 19 ----- 20
Db 2563 IIRQHDYGHWTDRSKLSLKEITNVQYVSCMNPAGSFTINPRLQRHESVFLSPGADA 2622
QY 21 ----- 20
Db 2623 LSIYSIILTOHLKLGNEPASLOKSIPPLIDLALFHOKIATTELPYGIKFHYIFNLRF 2682
QY 21 ----- 20
Db 2683 ANIFQILFSSVECYKSTWDLIRLYLHESNRKYRDKMVEKDFLEDKIQTEVULKTFDD 2742
QY 21 ----- 20
Db 2743 IEDPYEQTPNLCHPANGIGEPRKMPVQSWELLTQTLVEALENHNEVNTVMDLVLEFD 2802
QY 21 ----- 20
Db 2803 AMRHVCHNRILLESBPGNALLVGVGSGKOSLTRLAFTSSMDVFOITLRGQIODEFKM 2862
QY 21 ----- 20
Db 2863 DLASLCRAGVKNLNTVPLMTDAQVADERFLVINDLASGEIPDLYSDEVENIISNR 2922
QY 21 ----- 20
Db 2923 NEVKSQGLVDNRENCWKEFIDIRIROLKVTLCFSPVGNKLVRVSRKFPALVNCIAIHFH 2982
QY 21 ----- 28
Db 2983 EMPQALLESVSLRFLQNTGIEPTVKOSISKFMFVHTSVNOTSGSYLSNBORRYNTPK 3042
QY 29 ----- 28
Db 3043 SFLEFIRLYQSLHHRHREKCKTERLENGLLKLSHSAQYDDLKAKILAQAEVELKQNE 3102
QY 29 ----- 28
Db 3103 DADKLIQVVEVETDVSREKAMADEEQKVAVIMLEVQOKODEEDLAKAEPALTAQA 3162
QY 29 ----- 28
Db 3163 ALNLTNKTNLTELSFGSPPLAVSNVSAVMYLMAPRGVRPDRSRKAKAYTMAKVDGEL 3222
QY 29 ----- 28
Db 3223 DSLINPNKENIHENCLKAIIRPLYDDPEFNPEFVATKSYAAGLCSWVINIRFEVEPCDV 3282
QY 29 ----- 28
Db 3283 EPRKQALNKATADLTAAOEKLAIAKAKIAHLNENLAKLTARFEKATADKLCOQEAELYTA 3342
QY 29 ----- 28
Db 3343 VTIISLANRLVGLASENVRMADAVONFKQOEFTLCGDILLITAFISYGFETTKYROSLL 3402
QY 29 ----- 28
Db 3403 DRTWRPYLSQLKTIPIPVTPALDPLRLMLDDADVAAWONEGLPADRMSVENATILINCERW 3462
QY 29 ----- 28

```


Db 3463 PLAMDPOLOGIKWIKYGEDRLVTOIGOKYLOITIEQALBAGAVLIENTESIDPVLG 3522
 OY 29 -----GRF 31
 Db 3523 PLUGREVIKGRF 3535
 RESULT 5
 ZAN_PIG STANDARD; PRT; 2476 AA.
 ID ZAN_PIG 028983;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zonadhesin precursor.
 GN ZAN.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OC NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 823-830; 859-872; 883-890;
 RP 920-925; 960-967; 1235-1244; 1349-1354; 1518-1532; 1624-1656;
 RP 1658-1667; 1777-1795 AND 1914-1921.
 RC STRAIN=Meishan; TISSUE=Testis;
 RX MEDLINE=96064636; PubMed=7592795;
 RA Hardy D.M., Gathers D.L.;
 RT "A sperm membrane protein that binds in a species-specific manner to
 RT the egg extracellular matrix is homologous to von Willebrand
 RT factor".
 RL J. Biol. Chem. 270:26025-26028(1995).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: IN TESTIS, PRIMARILY IN HAPLOID SPERMATIDS.
 CC NOT IN LUNG, LIVER, HEART, SPLEEN, BRAIN, KIDNEY, EPIDIDYMIS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
 CC -1- DOMAIN: THE WVED DOMAINS 2 AND 3 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1- PTM: THE MAM DOMAINS AND THE MUCIN-LIKE DOMAINS ARE MISSING
 CC FROM THE ZONADHESIN THAT BINDS TO THE EGG EXTRACELLULAR MATRIX.
 CC PROCESSING MIGHT OCCUR DURING SPERM MATURATION AND/OR
 CC CAPACITATION.
 CC -1- SIMILARITY: CONTAINS 2 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4.5 WVED DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U40024; AAC48486.1; -.
 DR HSSP: P56682; IGV.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000998; MAM domain.
 DR InterPro: IPR002919; TIL-Cysrich.
 DR InterPro: IPR003328; TIL-Cysrich.
 DR InterPro: IPR001007; WVE_C.
 DR InterPro: IPR001846; WVE_D.
 DR Pfam: PF00094; vwd; 4.
 DR Pfam: PF00629; MAM; 2.

DR Pfam: PF01826; TIL; 5.
 DR Pfam: PF02345; TILA; 5.
 DR SMART: SM00181; EGF; 1.
 DR SMART: SM00137; MAM; 1.
 DR SMART: SM00214; VNC; 2.
 DR SMART: SM00216; VMD; 4.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 4.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS00600; MAM_2; 2.
 DR Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
 KW Repeat.
 FT SIGNAL 1 29
 FT CHAIN 30 2476
 FT DOMAIN 30 2418
 FT TRANSMEM 2419 2439
 FT DOMAIN 2440 2476
 FT DOMAIN 31 144
 FT DOMAIN 147 312
 FT DOMAIN 319 687
 FT DOMAIN 688 799
 FT DOMAIN 800 1184
 FT DOMAIN 1185 1573
 FT DOMAIN 1574 1968
 FT DOMAIN 1969 2370
 FT DOMAIN 2366 2402
 FT DISULFID 2370 2381
 FT DISULFID 2375 2380
 FT DISULFID 2392 2401
 FT CARBOHYD 109 109
 FT CARBOHYD 269 269
 FT CARBOHYD 755 755
 FT CARBOHYD 758 758
 FT CARBOHYD 833 833
 FT CARBOHYD 1134 1134
 FT CARBOHYD 1329 1329
 FT CARBOHYD 1448 1448
 FT CARBOHYD 1444 1544
 FT CARBOHYD 1544 1596
 FT CARBOHYD 1596 1654
 FT CARBOHYD 1654 1654
 FT CARBOHYD 1843 1843
 FT CARBOHYD 1965 1965
 FT CARBOHYD 2122 2122
 FT CARBOHYD 2165 2165
 FT CARBOHYD 2178 2178
 FT CARBOHYD 2329 2329
 FT CARBOHYD 2359 2359
 FT CONFLICT 823 823
 FT CONFLICT 923 923
 FT CONFLICT 965 965
 FT CONFLICT 1241 1241
 SQ SEQUENCE 2476 AA; 270364 MW; A13B690375A6548C CRC64;
 Query Match 39.7%; Score 62; DB 1; Length 2476;
 Best Local Similarity 1.6%; Pred. No. 4e+02;
 Matches 19; Conservative 2; Mismatches 7; Indels 1133; Gaps 5;
 OY 4 HX-----SMEKRT-----
 Db 762 HCHCRPSSMECQTKRGCHTHVYQQLKNGQYCHPYGSATCSYGDPHYLTFRDRFRNFMG 821
 OY 13 -----
 Db 822 KCTYIIAOPCGNLTEHPFRVLTKKKEKRGEGVSCLSKTYVTLPESTVTLKGRHTLVGGQ 881
 OY 13 -----
 Db 882 RYTLPAISRGVFLAPSGRHFVELQTAFLRYRWMDGDLFVSVPSTFSGKLGICGIDYG 941
 OY 13 -----
 Db 942 DSSNDNQKPDGSPADKDELGSWQTSDDAQCCENQVSPSCNTALQNTMSGPERCGQ 1001

```

QY 13 ----- 12
Db 1002 LVAPHGVFEACPLHRASSFEKCTEDMCFQGLQMLCAHMSALTENCODAGYVAPWR 1061
QY 13 ----- 12
Db 1062 GPQCPPLACPRNSRYTLCARLCPDTCSEFSGRACKRCVGECEODPGFVLSGLQCVSR 1121
QY 13 ----- 12
Db 1122 ECGCLDSTAGYVKEWFERPGKROLICBGNNTRCVLRCAOEFGQODIYGCHAO 1181
QY 13 ----- 12
Db 1182 GSATCVSGDPHYLTPDGAHLHFTGTCTYTLTKPCWLRLSENSELYSATNEFRGNLEAS 1241
QY 13 ----- 12
Db 1242 YVRAVOYQVFNLRISLIKGRKVTLDGRVALPLMPAQGRVSTSSGFFILYTDGLQVR 1301
QY 13 ----- 18
Db 1302 YDGDHLYEVTPSSYAGRLGCLGNNNNSLDILDPDKRPASSVRLGASWKINELSEP 1361
QY 19 ----- 18
Db 1362 GCFABGKRPKCLKEVADAMRKNDVLANPQSPFSCHRVAPQSSFSGLYGQCATKG 1421
QY 19 ----- 18
Db 1422 DTLFRLSLQAYASLACARAGQALTWNRGTCPCLKPSGSSYSTCANCPATCLSLNPSY 1481
QY 19 ----- 21
Db 1482 CPSTLPACBCECQKHLSTSCVPLSGCCCTTQGRSYHPVGSWTINDNSCLCTCSA 1541
QY 22 ----- 21
Db 1542 HNNISCRQASCKPQMCWPDGLRCRVAGKVCRIPTSHYVSGSYHAVRGNCTYVL 1601
QY 22 ----- 21
Db 1602 VKICHTMDLFPKISGENGRKQGPAPFLRYVYDIFNTLVTKODQVLLNGTRVSLP 1661
QY 22 ----- 26
Db 1662 ATTQIRGVAVSRDGYVTLTINIGVQKFDGRGFLEVEIPKAYYGRICGVGNFNDEED 1721
QY 27 ----- 26
Db 1722 ELMPSPDALDQVYVDSMRKDEIDPNCQEDDRKTEASEQEOPSSANCPRADERAOEC 1781
QY 27 ----- 26
Db 1782 QAAPQAPAMANCATRVVLSPYVRSCTHKLCEGGLNRAFCESLQAFGAQAOAGIKPPVW 1841
QY 27 ----- 26
Db 1842 RNSSTCPIDCSAHSVYTSVPSCLPSCODPBGCTGAGAPSTCEGGLCEGCVLYSRQC 1901
QY 27 ----- 31
Db 1902 VARSQCGCRDARGTFPLVGRF 1922

```

```

DE Protein ndvb.
GN NDVB OR R03286 OR SMC04382.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153914; PubMed=2154461;
RT Telpl L., Dylan T., Ditta G.S., Hellinski D.R., Stanfield S.W.;
RT "The ndvB locus of Rhizobium meliloti encodes a 319-kDa protein
RL involved in the production of beta-(1->2)-glucan.";
RL J. Biol. Chem. 265:2843-2851(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gauzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- FUNCTION: INVOLVED IN THE PRODUCTION OF BETA-(1,2)-GLUCAN.
CC IT IS INVOLVED NOT ONLY IN INVASION BUT ALSO IN BACTEROID
CC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: TO A.TUBERCULENS CHVB.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: J05219; AAA26305.1; ALT_INIT.
DR EMBL: AL591793; CAC47865.1; -.
KW PIR: A35548; A35548.
KM Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 411 431 POTENTIAL.
FT TRANSMEM 444 464 POTENTIAL.
FT TRANSMEM 810 830 POTENTIAL.
FT TRANSMEM 831 851 POTENTIAL.
FT TRANSMEM 959 979 POTENTIAL.
FT CONFLICT 47 47 T -> A (IN REF. 1).
FT CONFLICT 1187 1187 L -> F (IN REF. 1).
FT CONFLICT 2602 2602 D -> G (IN REF. 1).
SQ SEQUENCE 2832 AA; 315746 MW; 171EA89F03A936F2 CRC64;

Query Match 39.7%; Score 62; DB 1; Length 2832;
Best Local Similarity 0.8%; Pred. No. 5e+02;
Matches 14; Conservative 4; Mismatches 10; Indels 1694; Gaps 3;

```

```

QY 12 ----- 11
Db 382 KORLAEKRIKIGPSIFQHLIRSVKRLDMFAIAGPNILLTIAMIYVAVSPMDIPSGA 441
QY 12 ----- 11
Db 442 KLIMLILFALPASBEGAMGLFNTVTLFAKPSRLVGEYFLDIPEDARLLVYPCILAKRD 501
QY 12 ----- 11
Db 502 HDELVRLNEHYLANPGEIYFALLSPWADSKSEAPADTVLEYAKREIASLARYAY 561
QY 12 ----- 11
Db 562 DKTRFELLHRRRLYNEAGVMMGERKGLHELDLLRGDRDTSFLOGANMVEGYOY 621
QY 12 ----- 18
Db 622 VMTLSDPTLMDAVTKLVGKLYHPIINRPVVPRTQEVYTGSLIQPRVTSLTGSEAS 681
QY 19 ----- 18
Db 682 AFQRLFTINRGIDPYEYTVSDYQDIAGEGFTGKGLYHVAFAEALKSRIENAVLSHD 741
QY 19 ----- 18
Db 742 LLEGYARCALVTIDELVEDPPIRYEVEMSRQHRMARGDWOLLPYTFNPKNGLSMLGHWK 801
QY 19 ----- 18
Db 802 MYDNLRSLIPVAMLAASVWGMYMEPTPALIMQVLIFSLFVAPTLISGISIMPRRDI 861
QY 19 ----- 18
Db 862 VARAHLHTVLSDIRAANAQVALRIEIAHNAAMADAIYRSLYRFYSRKLMEWRTAAQ 921
QY 19 ----- 18
Db 922 VOSAGHSIGDYFRAMMTAPALALVSLAIAISDTGLPFLGFPALIMASPAVAMFVSQ 981
QY 19 ----- 18
Db 982 SAETEDQLVSEAEIEMKRIARTWRFEAFVTAQNFPPNDPOETPOPLAERTSPT 1041
QY 19 ----- 27
Db 1042 NIGVILLVMSARSGWIGFEETITRLQTIATIDRMKYRGHLFNMWYRTKGLSEMEPRY 1101
QY 28 ----- 27
Db 1102 VSSVDSGNLAGHLIAVSSKCREMAEAPSAHVQGNLDIGDVAAILKEALNELLPDDRKTIV 1161
QY 28 ----- 27
Db 1162 PLRLVEERLAFQNALAAVKKRELASTRVINTIAVLARDMKLTVINLDEHYRTVQSGEV 1221
QY 28 ----- 27
Db 1222 ATWAGSLVACEAHIAADVFDIGATIALRQRLVLKERARDAIFSMDSFLRPERRLIS 1281
QY 28 ----- 27
Db 1282 IGYRVANANELDACYDLASERLTSLEIAIAKGDLPTHEWTKLGRPIVIGARGALVMS 1341
QY 28 ----- 27
Db 1342 GSMFEYLPPLVMOROGSILNQTNINLVQBOINHGRLGTPWGISEAFANRDEHLYO 1401
QY 28 ----- 27
Db 1402 YTNFQVPTLGLKRGIGONAVIAPYASIIACMTDPSALANLARLREVGALGAYGHDAVD 1461
QY 28 ----- 27

```

```

Db 1462 FTPTREBEGOKCAVVRNYAHHGMVAAVANVENCOLREMFHADPVEIAELLQEK 1521
QY 28 ----- 27
Db 1522 PROIPVMAAKREPRALGQADLLREYRVVEDEPINODRETVLLSNGHSVMTLTAAGY 1581
QY 28 ----- 27
Db 1582 ARWNGSVTRWTPDPVEDRTGTFIFLBDYTGDMWSATAEPRAPGKTVTRFGDKAEF 1641
QY 28 ----- 27
Db 1642 VKTVGDLTSEVEICVATEHDAEGRVILLNTGDEREFEYTSYAEPLAMDADDSHPTE 1701
QY 28 ----- 27
Db 1702 SKMFLREISRHGDYIVWSRNKRSPGDDIEVAHLVTDNAGSERHQAETDRRRFLGQGR 1761
QY 28 ----- 27
Db 1762 TLAEAAFDPGATLSTGDTFTLDPITVSLRRVVRYPACKSVIFWTIAPDREGVDRAID 1821
QY 28 ----- 31
Db 1822 RYRHPETFNHLLHAWTRSOYOMRHVGTSEKASFOMLGRY 1863

```

RESULT 7

```

HGPA_HAEIN STANDARD; PRT; 999 AA.
ID HGPA_HAEIN AC 057408; P96344; 086244;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable hemoglobin and hemoglobin binding protein 4
DE precursor.
GN H11565/H11567.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Georgagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RT Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.,
RA Peterson J., Hickey E., Dodson R., Gwinn M.,
RA Submitted (MAY-1998) to the EMBL/GenBank/DBD databases.
RN [3]
RP CONCEPTUAL TRANSLATION.
RA Couderc E.;
RL Unpublished observations (AUG-2001).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (BY SIMILARITY).
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE COA

```

CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO
CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (By similarity).
CC -1 SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -1 CAUTION: THIS IS A CONCEPTUAL TRANSLATION; TWO FRAMESHIFTS WERE
CC INTRODUCED IN POSITION 289 AND IN THE REPEATS REGION TO MAXIMIZE
CC THE SIMILARITY WITH OTHER ORTHOLOGS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; U32831;	AAC23213.1;	ALT_SEQ.
DR	EMBL; U32831;	AAC23214.1;	ALT_SEQ.
DR	TIGR; H11565;	-	
DR	TIGR; H11567;	-	
DR	InterPro; IPR000531;	TonB_boxC.	
DR	Pfam; PF00593;	TonB_boxC; 1.	
DR	PROSITE; PS00430;	TON_DEPENDENT_REC_1; FALSE_NEG.	
DR	PROSITE; PS01156;	TON_DEPENDENT_REC_2; 1.	
KW	Outer membrane; Transport; TonB box; Multigene family; Signal; Receptor; Repeat; Complete proteome.		
RW	Receptor; Repeat; Complete proteome.		
FT	SIGNAL	1	24 POTENTIAL.
FT	CHAIN	25	999 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN BINDING PROTEIN 4.
FT	DOMAIN	26	49 1 X 4 AA TANDEM REPEATS OF P-T-N-Q.
FT	REPEAT	26	29 1.
FT	REPEAT	30	33 2.
FT	REPEAT	34	37 3.
FT	REPEAT	38	41 4.
FT	REPEAT	42	45 5.
FT	REPEAT	46	49 6.
FT	STATE	58	65 TONB_BOX.
FT	SITE	982	999 TONB_C-TERMINAL_BOX.
SO	SEQUENCE	999 AA:	114314 MW: DAECDAB700DA876 CAC64;

Query Match	39.18;	Score 61;	DB 1;	Length 999;
Best Local Similarity	2.28;	Pred. No. 99;		
Matches 16; Conservative	4;	Mismatches 11;	Indels 713;	Gaps

Qy	1	SRKXHSME	9
	1	: : : :	
Db	241	SKRHGELENYDKNGRDIQGREKADPYITTEKSTLVKFSFSPTENHRTVASDTYIQ	36
Qy	10		9
Db	301	HSRGDESYLVKTYIINKDEELRHNDLIRKKNVSTENTYVTVPMDLKISQQR	36
Qy	10	XRTPI-	15
	11	:	
Db	361	ITTRARTEDCDNEKCDSKNPLGLQLEKGVYDRNGDPVELKIVEDQGRHOVVK	42
Qy	16		15
Db	421	YNNFVSASTNDAAFVGQOLSPSEFWLDCIFNCDFRVKYQYNSQDESEVELNR	48
Qy	16		15
Db	481	TMEINRKFFATYESNNRYRDRYHMLIPNSGVLPLDYKERDLNTRKQINDLIRKATLE	54
Qy	16	NPWT-	20
	11	:	
Db	541	IENELSYGVYAKTKEMVNRKAGYGRNPTWMAERTLGLSLNGLRPTCKEDSSYNGLLCP	60
Qy	21		20

Db	601	RHEPSTSPLIPETTTKSLTYFADNKKILNMLSYDGLGYRDDIKKQPEIIPGVTRKIDDM	660
Oy	21	-----	20
Db	661	VRELFPVLPBPANGKDWGNPNVYTPQIINKNAENIAYIAQERPKKHSYSLGATFDPLNF	720
Oy	21	-----	20
Db	721	LRVQVYKSGGRPTPSDELTYTFKKHPDTILPNPMKPEAKNOELALTFHHNGFFSTN	780
Oy	21	-----	20
Db	781	VFQTYRQFIDLAYIGSNLNSVGGQAQADPQYQYQVNVNDRAKVKGEINSLNIGYF	840
Oy	21	-----	20
Db	841	FEKLDFNVSVYKFTYQGRGLDGNRPMAIOPKTSVIGLGDHKEQRFAGDIYVTHSAKK	900
Oy	21	-----	x 21
Db	901	AKDIYNNFYKIQGKSDASVRKMSDDYTLVDFVYIKPKVKNVTLQFGVYNTLDRRYLWES	960
Oy	22	XRGIRPVG-----RF 31	
Db	961	ARSIRPFGTSLNINCGTGAQINRF 984	

RESULT 8

OTOF_HUMAN	STANDARD:	PRT: 1997 AA.
ID	OTOF_HUMAN	09HC10: 09HC09: Q9Y650: 09HC08;
AC	15-JUN-2002 (Rel. 41, Created)	
DT	15-JUN-2002 (Rel. 41, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Otoferlin (Fer-1 like protein_2).	
GN	OTOF OR FER1L2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OX	NCBI_taxid=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND ALTERNATIVE SPLICING.	
RC	TISSUE=Brain;	
RX	MEDLINE=20395831; PubMed=10903124;	
RA	Yasunaga S., Grati M., Charpentoux S., Smith T.N., Friedman T.B.,	
RA	Iatwani A.K., Wilcox E.R., Petit C.;	
RT	"OtoF encodes multiple long and short isoforms: genetic evidence that	
RL	the long ones underlie recessive deafness DFNB9.";	
RL	Am. J. Hum. Genet. 67:591-600(2000).	
RN	[2]	
RP	SEQUENCE FROM N.A. (ISOFORM 4).	
RC	TISSUE=Fetal;	
RX	MEDLINE=99206603; PubMed=10192385;	
RA	Yasunaga S., Grati M., Cohen-Salmon M., El-Amraoui A., Mustapha M.,	
RA	Salen N., El-Zir E., Loiselet J., Petit C.;	
RT	"A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes	
RL	DFNB9, a nonsyndromic form of deafness.";	
RL	Nat. Genet. 21:363-369(1999).	
CC	-1- FUNCTION: Might be involved in the Ca(2+)-triggered synaptic	
CC	vesicle-plasma membrane fusion.	
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).	
CC	-1- ALTERNATIVE PRODUCTS: At least 4 isoforms: 1/long (shown here),	
CC	2/short-1, 3/short-2 and 4/short-3; are produced by alternative	
CC	splicing.	
CC	-1- TISSUE SPECIFICITY: Isoform 1 and isoform 3 are found in adult	
CC	brain. Isoform 2 is expressed in the fetus and in adult brain,	
CC	heart, placenta, skeletal muscle and kidney.	
CC	-1- DISEASE: Defects in OTOF are the cause of nonsyndromic autosomal	
CC	recessive deafness 9 (DFNB9).	
CC	-1- SIMILARITY: BELONGS TO THE FERLIN FAMILY.	
CC	-1- SIMILARITY: CONTAINS 4 C2 DOMAINS.	
CC	-----	

This SWISS-PROT entry is copyright It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; AF183185; AAG12991.1;	-
DR	EMBL; AF183186; AAG12992.1;	-
DR	EMBL; AF183187; AAG17468.1;	-
DR	EMBL; AF107403; AAD26117.1;	-
DR	HSSP; P04410; IAZ5.	
DR	Genew; HGNC:8515; OTOF.	
DR	MIM; 603681; -.	
DR	InterPro: IPRO000008; C2.	
DR	pfam; PF00168; C2; 8.	
DR	SMART; SM00239; C2; 6.	
DR	PROSITE; PS00499; C2_DOMAIN_1; 2.	
KW	PROSITE; PS50004; C2_DOMAIN_2; 4.	
DR	Transmembrane; Repeat; Alternative splicing; Deafness.	
FT	DOMAIN	1 1963
FT	TRANSLEM	1964 1984
FT	DOMAIN	1985 1997
FT	DOMAIN	241 338
FT	DOMAIN	404 514
FT	DOMAIN	947 1052
FT	DOMAIN	1479 1577
FT	DOMAIN	1303 1310
FT	DOMAIN	1314 1320
FT	DOMAIN	1965 1983
FT	VARSPLIC	1 747
FT	VARSPLIC	1245 1264
FT	VARSPLIC	1 690
FT	VARSPLIC	691 738
FT	VARSPIC	1943 1997
FT	CONFLICT	1088 1088
FT	SEQUENCE	1997 AA; 226735 MM; 39D10CB522063BAE CRC64;

Query Match	39.1%;	Score 61;	DB 1;	Length 1997;
Best Local Similarity	1.3%;	Pred. No. 3.5e+02;		
Matches 17;	Conservative	3;	Mismatches 10;	Indels 1241; Gaps

QY	2	RXHHSMEXRTPD-	-----	14
	:			
Db	183	KNNSHKEEPORPEPAVLEMEDLDHLAIRLGDGLDPDSVSLASVTALTITVNSKRSKPD1		242
QY	15	-----		14
Db	243	KMEPSAGRPMDYOVSTIVIEARQLVGLNMDPVYCYEGVDCKKITSMEKSTNCPYTNEYFV		302
QY	15	-----		14
Db	303	FDFHVSPPDVMEFKIIRISVISHSKNLLRSGTLVGSFKMDVGTIVYSOPREHQPHKKALLSDP		362
QY	15	-----		14
Db	363	DDISSGLKGYVCKDVAVVGKGDNIKTPHKANETDEDDIEGNLLRPEGVPPERQWARYRK		422
QY	15	-----		14
Db	423	IYRAEGLPRMNTSLMANVKKAFIGENKDLVDYVQVFFAGQKGKTSVQKSYEDLMEQV		482
QY	15	-----INPAW-----		19
		:		
Db	483	VFTDLFPLCKRKKVOIRDSCKVNDVAIGTHFIDLRKISNDGDKGFLPTLGPANVNNYS		542
QY	20	-----		19

Db	543	TRNTLLDEHODLNEGSGEVSEPARLLGLAVEIVDTSNPELTSSTVEYVQAPRIS	602
QY	20	-----	19
Db	603	CAGKEEFFLEGAFLEASMSIDRRNDKPTFEVTIGNYGENVDGLSRQRPAPKEPGE	662
QY	20	-----	19
Db	663	EEVDLIQNASDDEAGDGLASVSTPPMRPOVTDNRNPHLEPLERKPCIIYKSNWPDQR	722
QY	20	-----	19
Db	723	RRLYANIMDIADKLEEGINDIOEMIKTEKSYPERRLRGVLEELSCGCCREFLSLADKQ	782
QY	20	-----	19
Db	783	GHSSTRILDRERLKSCRELENNGOQARMLRAQVKHRYRDLRLCQNFLOKRLADEP	842
QY	20	-----	19
Db	843	QHSIPDIFIMMSNNKRVATARYPSKDLLESTVEEETGCDKAVTLPILKLRKNGFSA	902
QY	20	-----	19
Db	903	GWTVOAKVELYLMGLSKQRKEFLCGPGQGEVKAQGLGHAFPPVSLVYTKKQAFQL	962
QY	20	----YXXR-----	23
Db	963	RAHMTQASLFPADSSGLSDPFARVEFFINOSQTEVYLNELTLCPTWDMLVFDNLELYGEA	1022
QY	24	-----	23
Db	1023	HELADDPPIYIEIYDQDSMGKADPFMGRTFAKPLVMADGACPPRFPQLEYQIYHGN	1082
QY	24	-----GIRP-----	27
Db	1083	ATAGDPLAPELLQIGPGRADLPPIINGPYVDVDRGEIMPEVPGIRPVLSKYREVLFWGL	1142
QY	28	-----	27
Db	1143	RDLKRVLAQYDRPRDIECAGKGVOSSLIHNYKKPNFNTLVKWEYDLPENELLAPPL	1202
QY	28	-----	27
Db	1203	NIRVYDQARGRYTVGSHAVSSLRNFYIRPPDRSAPSMNTTVRLRRCRYLCNGSSSH	1262
QY	28	-----	27
Db	1263	STGEVVTMEPEVPKIKLETMTKLDATSEAVVVDVAEEKEKKKKKGTAEPPEEPEPD	1322
QY	28	-----	27
Db	1323	ESMLDMSSKIFASIDIMKEQLROQEPSGIDLEEKEVDNTEGLKSGMKGEKARAAREK	1382
QY	28	-----	27
Db	1383	KKTKQSSGSGGSGAPKKPKIDELKVPKLESEFPDFEDMLHTFNNLRCKTGDDDEG	1442
QY	28	-----VGRF 31 	31
Db	1443	STEEERIVGRF 1453	1453
RESULT 9			
OTOF_MOUSE			
AC	Q9ESF1; Q9ESF2;	STANDARD;	PRT; 1997 AA.
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	OTofeRlin (fer-1 like protein 2).		
GN	OTOF OR FER12.		

OS Mus musculus (Mouse).
OC Euryotia; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND ALTERNATIVE SPLICING.
RC TISSUE=Brain, and cochlea;
RX MEDLINE=2035831; PubMed=10903124;
RA Yasunaga S., Grati M., Chardenoux S., Smith T.N., Friedman T.B.,
RA Lalwani A.K., Wilcox E.R., Petit C.;
RT "ORF clones multiple long and short isoforms: genetic evidence that
RT the long ones underlie recessive deafness DFNB9.";
RL Am. J. Hum. Genet. 67:591-600(2000).
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=99206603; PubMed=10192385;
RA Yasunaga S., Grati M., Cohen-Salmon M., El-Amraoui A., Mustapha M.,
RA Salem N., El-zir E., Lolselot J., Petit C.;
RT "A mutation in ORF, encoding otoferlin, a FER-1-like protein, causes
RT DFNB9, a nonsyndromic form of deafness.";
RL Nat. Genet. 21:363-369(1999).
CC -1- FUNCTION: Might be involved in the Ca(2+)-triggered synaptic
CC vesicle-plasma membrane fusion.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (by similarity).
CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms; 1 (shown here) and 2;
CC are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Strongly expressed in brain and inner ear. In
CC the inner ear, it is mainly expressed in the cochlear IHC and
CC vestibular type I sensory hair cells. Weakly expressed in eye,
CC heart, skeletal muscle, liver, kidney, lung and testis.
CC -1- SIMILARITY: BELONGS TO THE FERLIN FAMILY.
CC -1- SIMILARITY: CONTAINS 4 C2 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/
CC or send an email to licensing@isb-sib.ch](http://www.isb-sib.ch/announce/or-send-an-email-to-licensing@isb-sib.ch)).
CC -----
DR EMBL; AF181813; AAG12989.1; -
DR EMBL; AF181814; AAG12990.1; -
DR HSSP; P04410; 1A25.
DR MGD; MGI:1891247; Ocof.
DR InterPro; IPR000008; C2.
DR Pfam; PF00168; C2; 8.
DR PRINTS; PRO0360; C2DOMAIN.
DR SMART; SMO0239; C2; 6.
DR PROSITE; PS00499; C2_DOMAIN_1; 2.
DR PROSITE; PS50004; C2_DOMAIN_2; 4.
RW Transmembrane; Repeat; Alternative splicing.
FT DOMAIN 1 1963 CYTOPLASMIC (POTENTIAL).
FT TRANSMM 1964 1984 POTENTIAL.
FT DOMAIN 1985 1997 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 240 337 C2 DOMAIN 1.
FT DOMAIN 403 513 C2 DOMAIN 2.
FT DOMAIN 946 1051 C2 DOMAIN 3.
FT DOMAIN 1479 1577 C2 DOMAIN 4.
FT DOMAIN 1305 1310 POLY-TYS.
FT DOMAIN 1314 1320 POLY-GLU.
FT DOMAIN 1970 1978 POLY-LEU.
FT VASPLIC 169 169 R -> SKREETGKGRGDEHK (IN ISOFORM 2).
FT VASPLIC 1244 1263 MISSING (IN ISOFORM 2).
FT VASPLIC 1943 1997 SEIWEIINDKSAAREYIMHWYKLLKFLLELLLELALFL
FT VASPLIC 1943 1997 YSLGGLAKTKGA -> AFWPLNPLKSLIKYILCTRYKWL
FT VASPLIC 1943 1997 IYKIVLALIGLMLALFLYSLPDGIVWKLUSA (IN
FT ISOFORM 2).
FT SEQUENCE 1997 AA; 227031 MW; 7B9CCA918F79D4D1 CRC64;
FT

Query Match 39.1%; Score 61; DB 1; Length 1997;
Best Local Similarity 1.3%; Pred. No. 3, 5e+02;
Matches 18; Conservative 1; Mismatches 6; Indels 1316; Gaps 4;

OY	7	SMEKR-----	7	11
Db	113	SMEVRYQATDGTVPMDGDPLDESLQEBKDSQETDGLLPSRSRSTRISGEKSPRAGR	172	
OY	12	-----	11	
Db	173	SVFSAMKIGKTRSHKEEPORQDEPAVLEMEDDLHLA1QLGDLDPDSVLSVTLATSNV	232	
OY	12	-----	11	
Db	233	SNKRSKPDIMKPSAGNPRMDYQVSTVIEARQLVGLMDPVYCVCEYGDCKTYSKKESTN	292	
OY	12	-----	11	
Db	293	CPYINEXYFDEPHVSPDWMEDK1IKISVLSHKNLBSGTLVGSPKMDYGYYSQDEHQH	352	
OY	12	-----	11	
Db	353	HKMAILSDPDDISAGLKGYYKCDVAVYGKDN1KTPHKANETDEDD1EGNLLPEGYPE	412	
OY	12	-----	11	
Db	413	ROMARFYVKIYRAGELPRMNTSLMANYKKAFIGENKDLVPYQVFPAGQKGTYSOKSS	472	
OY	12	-----	16	
Db	473	YEP1MNEQVYFTDLEPPLCKRMKVQ1RNSDPKNDVA1GTHFDLRKISNDGDKCPLTLG	532	
OY	17	PAM-----	19	
Db	533	PAMVNMGTSTNRTLLDEHODLNEG1GEGVSFRARLMLGLAVE11DTSNPELTSTEQV	592	
OY	20	-----	19	
Db	593	EQATPVSEBCTGRABEEFFLGAFLEASMDRKNKGDKPITFEVTIGNYGENYDGMGRPLRP	652	
OY	20	-----	19	
Db	653	RPRKEPDDEEVDL1ONSSDDEGDEAGDLASVSTPRMPQ1TDRNYFHLPLYERKPC1Y	712	
OY	20	-----	19	
Db	713	IKSMWPDQRRLYMANIMDIADKLEGLNDVQEM1KTEKSPERRLRGVLELSCGHR	772	
OY	20	-----	19	
Db	773	FL1SDKDGKSSRTLRDLRERLSCMRLEMSGQAKSLRAQVRRHTYADKLRLCONFLQ	832	
OY	20	-----	19	
Db	833	KLRFLADEPQHS1PDVFIWMMSNKK1AARVPSKDLFSIYEBELGCKADKVKTLFLKL	892	
OY	20	-----	19	
Db	893	PKRGFGSAGTWQAKLELYLMLG1SKQRKDFLCBPCGFEEVKAAG1GLSHSP1SLV	952	
OY	20	-----	23	
		---YXXR---		
Db	953	YTKRQAFQ1LRAHMTQARS1FLADSSGLSDPFARVFFINOSCTEYV1NETLCTPTMDQ1VF	1012	
OY	24	-----	23	
Db	1013	DNLELYGEAHELNDPPI1YIETIYDQDSMGKADFMGRTPAKPLVYMADEAYCOPRPPQ1L	1072	
OY	24	-----	27	
		-----G1RP-----		
Db	1073	EYQ1YRGSATAGD1LAFAELLQ1GSPGKADLP1NGPYDMRGCP1MPVGI1REV1SKY	1132	
OY	28	-----	27	
Db	1133	RVEVLEFGLDR1KRVN1AQVDRPRVD1ECAGKGVSS1L1HNTKKPNPNT1LYKFEVDLP	1192	

QY 28 ----- 27
 Db 1193 ENELLHPPLNIRVYDCRACFRGTYLVGSHAVSLNRFYRPPDSAPAMWTTVLLRQCHR 1252
 QY 28 ----- 27
 Db 1253 LRNGSPSPRPGEVVSEPEPEVKLETWVKLDANSDAVYKVADEKERKKKKKKGP 1312
 QY 28 ----- 27
 Db 1313 SEEPDEPEDESMMDWMSKYFASIDYMKEQLRHETSCTDLEKEMESAELGKPMKSK 1372
 QY 28 ----- 27
 Db 1373 EKSHAAEKKKKKNSPGSGSEAPKKAKIDELKVYPKEISEPDSFEDMLHFFENLL 1432
 QY 28 ----- 1453
 Db 1433 RGTGDEDEGSTEEERLVGRF 1453
 RESULT 10
 FAS2_DROME STANDARD; PRT; 873 AA.
 AC P34082: P34083:
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Fasciclin II precursor (FAS II).
 GN FAS2 OR EG:EG0007.3.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RX [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=Canton-S;
 RX MEDLINE=92005695; PubMed=1913818;
 RA Greeningloh G., Rehm E.J., Goodman C.S.;
 RT "Genetic analysis of growth cone guidance in Drosophila: fasciclin II
 RL functions as a neuronal recognition molecule.";
 RN Cell 67:45-57(1991).
 RC SEQUENCE OF 22-873 FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
 RA Dreano S., Gloux S., Leleure V., Mottier S., Gallibert F., Borkova D.,
 RA Minana B., Katos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA Papayianakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourikotl F.,
 RA Beinhart N., Dove G., Schaefer U., Jaackle H., Buchner A.,
 RA Callister D.M., Campbell L.A., Darlamsou A., Henderson N.S.,
 RA McMillan P.J., Sales C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of D.
 RT melanogaster.";
 RL Science 287:2220-2222(2000).
 CC -1- FUNCTION: Neuronal recognition molecule for the MP1 axon pathway,
 CC pathway recognition for axons during the development of nerve
 CC fascicles.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1);
 CC attached to the membrane by a GPI-anchor (isoform 2).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1/Membrane-linked (shown here)
 CC and 2/Phosphatidyl-inositol-linked, are produced by alternative
 CC splicing.
 CC -1- TISSUE SPECIFICITY: In embryos, both isoforms are initially
 CC expressed on the surface of the axons in the MP1 pathway and later
 CC on several other longitudinal axon fascicles.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: M7165; AAA28527.1; -
 DR EMBL: M7166; AAA28528.1; -
 DR EMBL: AL033125; CAA21825.1; -
 DR EMBL: AL033125; CAA21826.1; -
 DR PIR: A41054; A41054.
 DR PIR: B41054; B41054.
 DR FLYBASE: FBgn0000635; Fas2.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003598; IG_C2.
 DR InterPro: IPR003600; IG_Like.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF00047; ig; 5.
 DR SMART: SM00060; FN3; 2.
 DR SMART: SM00410; IG_Like; 2.
 DR SMART: SM00408; IGC2; 3.
 DR Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
 KW Immunoglobulin domain; Transmembrane; GPI-anchor; Signal;
 KW Neurogenesis.
 FT SIGNAL 1 28
 FT CHAIN 29 873
 FT DOMAIN 29 751
 FT TRANSMEM 752 769
 FT DOMAIN 770 873
 FT DOMAIN 47 123
 FT DOMAIN 152 214
 FT DOMAIN 244 309
 FT DOMAIN 336 414
 FT DOMAIN 444 511
 FT DOMAIN 544 619
 FT DOMAIN 648 705
 FT DISULFID 54 116
 FT DISULFID 159 207
 FT DISULFID 251 302
 FT DISULFID 343 407
 FT DISULFID 451 504
 FT CARBOHYD 74 74
 FT CARBOHYD 250 250
 FT CARBOHYD 330 330
 FT CARBOHYD 448 448
 FT CARBOHYD 458 458
 FT CARBOHYD 576 576
 FT CARBOHYD 738 811
 FT VARSPLIC 812 873
 FT CONFLICT 804 804
 FT SEQUENCE 873 AA; 96926 MW; E48F048DCB62AC9 CRC64;
 Query Match 38.5%; Score 60; DB 1; Length 873;
 Best Local Similarity 4.7%; Pred. No. 1e+02;
 Matches 16; Conservative 2; Mismatches 8; Indels 318; Gaps 4;
 QY 6 HSM---EKRTD-----INPAM---19
 Db 515 HDMQKEARVDEVSSEARPSQLTATMTDIRGSPTELGPIILASYQYKEALNPDWSTA 574
 QY 20 -----YXRGIRPV-----11
 Db 575 YNRSMSPSPYIVSELRLQRTETSPFARARNOYGLGNMGVNOQOSTPRRSAPPEPKPLHNP 634

OY 29 ----- 28
 DB 635 VOHDKPEPVVSPYSDHFLRMGVPADNGEPIDRQIKYCGVKISGTWLENSCMYVE 694
 OY 29 ----- 28
 DB 695 VMETTSFEMTQLVGNTYYRIELKANAIGYSPASIIKRTGRIDVIOVAEROVSSAAI 754
 OY 29 ----- 28
 DB 755 VGAIAGVLLLFVVDLCCTIVHGVAMATMCKRAKRSPELDEAKIGSGQLYKEPPPS 814
 OY 29 ----- GRF 31
 DB 815 PLPLPPVKGSGPMSTPLEDERPLRTPTGSIKONSTIEFDRGF 858
 RESULT 11
 ID UVRA_METTH STANDARD; PRT; 962 AA.
 AC 026543;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Exonuclease ABC subunit A.
 GN UVRA OR MTH443.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea: Euryarchaeota: Methanobacteria: Methanobacteriales;
 CC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 Altridge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 Harrison D., Hoang L., Keagle P., Lumm W., Potlter B., Qiu D.,
 Spadafino R., Viore R., Wang Y., Wierzbowski J., Gibson R.,
 Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
 Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RA "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT delah: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7153(1997).
 CC -1- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
 CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
 CC AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
 CC STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
 CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRA AND UVRC.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVRA SUBFAMILY.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVRA SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE000828; AAB84949.1.
 DR InterPro: IPR003439; ABC_transportr.
 DR InterPro: IPR004602; UVRA.
 DR Pfam: PF000005; ABC_tran.2.
 DR Prodom: PD000006; ABC_transportr.1.
 DR TIGRfam: TIGR00630; uvra.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER.2.
 KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
 KM DNA-binding; zinc-finger; Complete proteome.
 FT NP_BIND 38 45 ATP (POTENTIAL).
 FT NP_BIND 649 656 ATP (POTENTIAL).
 FT ZN_FING 748 774 C4-TYPE.

SQ SEQUENCE 962 AA; 108395 MM; 2C0EF7FC41CCD060 CRC64;
 Query Match 38.5%; Score 60; DB 1; Length 962;
 Best Local Similarity 3.1%; Pred. No. 1.2e+02;
 Matches 17; Conservative 2; Mismatches 10; Indels 511; Gaps 3;
 OY 2 2RXHSGME----- 9
 DB 201 RNRKHSIDVYDLVVRDTEFRKRLADSVEVATLQEGSTVAVPTMTGEEIRYSEHRA 260
 OY 10 ----- 9
 DB 261 CPGTAGINEFEISPRMSESPHGCPCNGCLGSKLEIDPLVVPERSINEGAIVPWS 320
 OY 10 ----- 9
 DB 321 KSGKRDYHYQMLRAVAEHGFSLDTPFDLDEHRRALILYGTDEKIQVFORKNRTYRV 380
 OY 10 ----- 9
 DB 381 NRRFEGVTPMERIYMETKSNMYRTYIGRMSNHACPGSGSRLPESLVTINGRSIHD 440
 OY 10 ----- 9
 DB 441 VVENSIREAHEPDSLKTEREYIAREVAKETRELRPLIDVGLDYTLSSSGTLSGG 500
 OY 10 ----- 9
 DB 501 EAQRIRLATQIGSGIVGLYLDEPSIGLQHRNRLIETIKRLRLGNTLIVHEDEET 560
 OY 10 ----- 9
 DB 561 ILSADHYVDIGPGAGEHGGCVAAEGTPEIEMEDPSLTGAYLSGRETIPLPEVRRRPSGR 620
 OY 10 ----- 9
 DB 621 YLYVGAENNLAEIDVRIPLGLFTCVTGVSGSGKSTLVNDILRYVERLHKHNAQR 680
 OY 10 ----- XRTPDINPAMYX----- XRGIRPVGRF 31
 DB 681 HTDIEGLQHDKYVMIDSPIGRTSPSNPATYGVFTHIRELFAQTPKARKGYRP-GRF 739
 RESULT 12
 ID HRX_MOUSE STANDARD; PRT; 3866 AA.
 AC P55200;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zinc finger protein HRX (ALL-1) (Fragment).
 GN MLR OR HRX OR ALL1.
 OS Mus musculus (Mouse).
 OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi;
 CC Mammalia: Eutheria: Rodentia: Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN-C57BL/6J, and C57BL/6 X CBA; TISSUE-Spleen, and Lung;
 RX MEDLINE=93317679; PubMed=8327517;
 RA Ma Q., Alder H., Nelson K.K., Chatterjee D., Gu Y., Nakamura T.,
 Canaan E., Croce C.M., Sircusa L.D., Buchberg A.M.;
 RA "Analysis of the murine All-1 gene reveals conserved domains with
 RT human ALL-1 and identifies a motif shared with DNA
 RT methyltransferases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC MAY REGULATE GENES INVOLVED IN SKELETAL FORMATION DURING
 CC EMBRYONIC DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.


```

CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L17069; AAA62593.1; -.
DR MGI: MGI:96995; M1.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003889; FYRICH.C.
DR InterPro: IPR003888; FYRICH.N.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR001214; SET.
DR InterPro: IPR002857; Zn1_CXXC.
DR InterPro: IPR001965; Zn1_PHD.
DR Pfam: PF00628; PHD; 3.
DR Pfam: PF00856; SET; 1.
DR Pfam: PF02008; 2F-CXXC; 1.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50280; SET; 1.
DR PROSITE: PS01359; 2F_PHD_1; 3.
DR PROSITE: PS50016; 2F_PHD_2; 3.
KW DNA-binding; Nuclear protein; Zinc-finger; Metal-binding;
KW Transcription regulation; Alternative splicing; Polymorphism.
FT NON_TER 1
FT DNA_BIND 67 78 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 115 125 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 199 207 A.T HOOK (BY SIMILARITY).
FT ZN_FING 1044 1091 CXXC-TYPE 1.
FT ZN_FING 1330 1381 PHD-TYPE 1.
FT ZN_FING 1383 1432 PHD-TYPE 2.
FT ZN_FING 1465 1529 PHD-TYPE 3.
FT ZN_FING 1605 1650 BROMODOMAIN (DIVERGENT).
FT DOMAIN 3737 3846 SET.
FT DOMAIN 35 41 POLY-GLY.
FT DOMAIN 459 469 PRO-RICH.
FT DOMAIN 1231 1238 POLY-PRO.
FT DOMAIN 3533 3536 POLY-GLU.
FT DOMAIN 3693 3697 POLY-GLU.
FT VASPLIC 1503 1505 MISSING (IN ISOFORM 2).
FT VARIANT 1497 1497 K -> T.
SQ SEQUENCE 3866 AA; 420976 MW; ADFC55E14E806F1D CRC64;

Query Match 38.5%; Score 60; DB 1; Length 3866;
Best Local Similarity 0.5%; Pred. No. 1.4e+03;
Matches 15; Conservative 6; Mismatches 9; Indels 2968; Gaps 4;
QY 2 RXHSHMEKRT----- 12
DB 647 RSPSHSMRTSGRLSTSELSPTPPSSVSSLSIPVSLAALNPFTTFPSHSLTOSGD 706
QY 13 ----- 12
DB 707 STEKNQBARQTSAPAEFFSSNPALFPWTFPGSGTEKGRKKDPAPELSKDRADKSYE 766
QY 13 ----- 12
DB 767 KDSREDREREKENKRESREKRRKGSIDIOSSALYPVGRVSEKAVAGEVGTSSAKK 826
QY 13 ----- 12

```

```

DB 827 ATGRKSSSLDGDVAVPTLGDITAVKAKILIKGRGNLEKNNLDIGPAAPSLERERP 886
QY 13 ----- 12
DB 887 CLASPSSYVKHSTISIGSLMAQADKLPMTDKRVASILKAKAOLCKIEKSLKOTDOP 946
QY 13 ----- 12
DB 947 KAQGESDSETSYRGPRIKHYVRAAVALGRKRAVPPDMLTSLAPWEEEREKILSMG 1006
QY 13 ----- 12
DB 1007 NDKSSVAGEDAEPLAPPIKPYTRNKAPQEPYKGRSRGCGCPGCPEDCGI 1066
QY 13 ----- 12
DB 1067 CTNCLDKPKFGGRNKKQCKMKRCONLWMPKASIQKOTKAVKKKKKTEKKEK 1126
QY 13 ----- 12
DB 1127 ESTSVKSPLEPAQKAPPPREAPPKSSSEPPPRKVEEKSEGGAPAPAPAEPKQYS 1186
QY 13 ----- 12
DB 1187 APASRKSSQVOPAAVVPPOPPSTAPQKKAPAVSEPKKQPPPEPGPSQKQKV 1246
QY 13 ----- 12
DB 1247 APLPSITVKKPKDKKEPPPVSKQENAGTINIPLNSGISSKOKIPADGVHRIYDFKE 1306
QY 13 ----- 12
DB 1307 DCEAENVEMGGILTSVPTPRVVCFLSSSEHYEFVYCYCCPEFHKCLEENERPL 1366
QY 13 ----- 12
DB 1367 EDQLENNCCRCKFCVCHQGHQATKQLLECNCRNSYHPECIGPNYPTRTKKKRWIC 1426
QY 13 ----- 12
DB 1427 TKCVKSCSGSTTPGKQMDQWMSHDSLCHDCAKLPKAKNFCPLCDKYDDDESKMNO 1486
QY 13 ----- 12
DB 1487 CGKCDRWVHKSCELSGTEDEMEILSNLEPSVAITCVNCTERHPPERLALKEIQASL 1546
QY 13 ----- 17
DB 1547 KOVLTALLNSRTTSHLRVQAAPKPDLPNTEESIPSRSSPGDPPLVTEVSKODEOQ 1606
QY 18 ----- 17
DB 1607 PLDEGKRRKMDQGSYVYLEFSDDIYKIIQAAINSDDGQPEIKKANSYKSFPIQMER 1666
QY 18 --ANYXXR----- 23
DB 1667 VFPEFSYKSRFEWPNKVNNSGMLPNAVLPPSLDHNVAQWQRESSHTEDQPLMKKII 1726
QY 24 ----- 23
DB 1727 PAPKPGGEPDSPTPLHPPTPLTSDRSREDSPELNPPIIDNRKQALCIMYGDGSA 1786
QY 24 ----- 23
DB 1787 NDAGRLIYIGONEMTHVNCALMSAEVPEDDGSLKNVHAIVINGKQLRCGFQCPGATVG 1846
QY 24 ----- 23
DB 1847 CCLTCTSNYHFMCSRAKNCVFLDDKKVYQCQRHDLIKGEVPENGEVEYRRAVDFDEGI 1906
QY 24 ----- 23

```

```

Db 1907 SLRRKFLNGLEPENIHMMIGMTIDCLGILNDLSDCEDKLEPIGYOCSRYVWSTTDARKR 1966
QY 24 ----- 23
Db 1967 CVYICKIMECRPYVEPDINSTVEHDNRTIAHSPSSFIDASCDSOITAILSPSPDR 2026
QY 24 ----- 23
Db 2027 PHSOTSGCYHYHISKVPRITPSYPTQSRPGCRPLPSAGSPPTTHEIYVGDPLISS 2086
QY 24 GIRPVG----- 29
Db 2087 GLRSTIGSRHSTSLSLPKRLIMSPYRGSAVSRSSVSSPSLGTATDPEASAKASDR 2146
QY 30 ----- 29
Db 2147 GGLSSANLGHSAHPSSSSQRTVGSKTSHLDGSSPEVKRCALDLVPKGLVKGEKN 2206
QY 30 ----- 29
Db 2207 RTSSSKSTDSGAHSTAYGPILKPLPOVHNATPGELINISKIGFAEPSTVPSSKQTVSYP 2266
QY 30 ----- 29
Db 2267 QLHLRGQRSDQHMDSQSVKPSPNEDGEIKTLKLPGMGRPSILHEHIGSSRDROR 2326
QY 30 ----- 29
Db 2327 GKSSKETCKEKHSSKSTLEPGQYTTGEGNLKPEFADEVLTTPGLQORPCNNVSEKIG 2386
QY 30 ----- 29
Db 2387 DKVLPLSGVPKGSTOYEGSKELQAPRKCSVKYPLKMEGENSKNTOKESGSPAH 2446
QY 30 ----- 29
Db 2447 ESVCPAEVSASRSPGACPGVQPSPNNTLSODPOSNNYQNLPEODRNLMLPDGPKQEDG 2506
QY 30 ----- 29
Db 2507 SEKRRYPRRSARARSNMFFGLTPLYGVASYEEDIPFYSNSTGKRGKRSAGQVDGAD 2566
QY 30 ----- 29
Db 2567 LSTDEDDLYYNTFRVVISGGGERLASHNLFREEQCCLPKISOLDGVDDGESDTSY 2626
QY 30 ----- 29
Db 2627 TATSRKSSQITKRNGKENGENTENKIDRPEDAGEKEHYIKSAVGHKNEPKLDNCHSVSRV 2686
QY 30 ----- 29
Db 2687 AQGDOSLEAQLSSLESSRNVHTSTPSDKNLDYNAELKSDSDNNNSDDCGNLLPSDIM 2746
QY 30 ----- 29
Db 2747 DEVLKNTPSMQALGESPESSSELLTLEGILGDSNREKDIGLEEVSSQOLPATEPVSS 2806
QY 30 ----- 29
Db 2807 VSSSISAEQFELPLEPLSDLSVLTTSPIYPSQNPRLAVISDSGKRVTTTEKSVASS 2866
QY 30 ----- 29
Db 2867 EGDPAALSPGVDPAPEGHMTDFHIQGHMADHISPPCGSVEGHNQSDLTRNSGTPG 2926
QY 30 ----- 29
Db 2927 LQVPSPTVPQONOKYVPSSTDSGSPQISNAVAQTTPHLKPATEKLIYVNMQMPLYV 2986
QY 30 ----- 29
Db 2987 LQTLPGVYQKIQLTSPVSVSTPSVMTNTSVLGPMSGGLTTLTGILNPSLPPSPSLPPAS 3046

```

```

QY 30 ----- 29
Db 3047 KGLLSVPHHQHLHSFPAQAASSFPPISSPPSGLLIGVQPPDPQLLGSANQRTDLTT 3106
QY 30 ----- 29
Db 3107 VAPSSGLKRRPISRLHTRKNNKLAPSSAPSNIAPSDVSNMTLINFTPSQLSNHPSLD 3166
QY 30 ----- 29
Db 3167 LGLSHPSSHRYVNIITKRSKSGIMFEQAPLLPPOSVGCTAATAAGSSITISDTSHLTSG 3226
QY 30 ----- 29
Db 3227 PVSALASGSSVLNVSMQTAAPTSSTSVPGHVTLANQRLGTDPDGSISHLIKASHQS 3286
QY 30 ----- 29
Db 3287 LGIQDPVALPSSSGMFPOLGTSQTPSAAMTAASSICVLPSSQTAGMTAASPGEABEH 3346
QY 30 ----- 29
Db 3347 YKLQGNQLLAGKTGTLTISQDRDPDAPGTQPSLFTQTAEPNGVSLQNKTLPSARPA 3406
QY 30 ----- 29
Db 3407 SSASPGSSPSGQSGSSVPGPTKPKAKRIQLPDKGSVKKHVSHLRTSSSAHLP 3466
QY 30 ----- 29
Db 3467 RDTDPAPQPSVTRTPRANREQDAAAGVEQPSQEGGAPGAPVALPEVQATQNPANQEN 3526
QY 30 ----- 29
Db 3527 AEPKAMEEESGFSPLMLMLQEQKREKSTERRKPKGLVFEISSDDGFOICAESIDA 3586
QY 30 -----RF 31
Db 3587 WKSLTDKVOEARSNARLKLQLSFAGVNGLEMLGLHDVAVFLEIQLAGAKHCENYKFR 3644

```

```

RESULT 13
HRX_HUMAN
ID HRX_HUMAN STANDARD; PRT; 3969 AA.
AC 003164; 014845; 016364; 013743; 013744; 09UWA3;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein HRX (ALU-1) (Trithorax-like protein).
GN MLR OR HRX OR ALU1 OR TRX1 OR HTRX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93046667; PubMed=1423624;
RA Tkachuk D.C., Kohler S., Cleary M.L.;
RA Marschalek R.;
RT "Exon/intron structure of the human ALU-1 (MLT) gene involved in
RT translocations to chromosomal region 11q23 and acute leukaemias.";
RL Br. J. Haematol. 93:966-972(1996).
RN [3]
RP SEQUENCE OF 1-1909 FROM N.A.
RX MEDLINE=93390935; PubMed=8378076;

```

RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
 RA Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.,
 RA "Two distinct portions of t(11;19)(p21;q23) are involved in t(11;19)
 RA leukemia.";
 RT Oncogene 8:2617-2625(1993).
 RN [4]
 RP SEQUENCE OF 1317-2328 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93265134; PubMed=1303259;
 RA Djabali M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.;
 RA "A trithorax-like gene is interrupted by chromosome 11q23
 RA translocations in acute leukaemias.";
 RT Nat. Genet. 2:113-118(1992).
 RL [5]
 RP SEQUENCE OF 1251-1538 FROM N.A.
 RX MEDLINE=94215165; PubMed=8162575;
 RA Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Canani O.,
 RA Saito H., Croce C.M., Canani E.;
 RA "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
 RA involved in acute leukemia.";
 RT Cancer Res. 54:2326-2330(1994).
 RN [6]
 RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
 RX MEDLINE=95322025; PubMed=7598802;
 RA Mbongkollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
 RA Rowley J.D., Diaz M.O.;
 RA "The human MLL gene: nucleotide sequence, homology to the Drosophila
 RA trithorax-like gene, and alternative splicing.";
 RT DNA Cell Biol. 14:475-483(1995).
 RN [7]
 RP SEQUENCE OF 1212-1603 FROM N.A.
 RX MEDLINE=95315013; PubMed=7794749;
 RA Marschallner R., Grell J., Lochner K., Nilson I., Slegler G.,
 RA Zwickroner I., Beck J.D., Fey G.H.;
 RA "Molecular analysis of the chromosomal breakpoint and fusion
 RA transcripts in the acute lymphoblastic SEM cell line with chromosomal
 RA translocation t(4;11).";
 RT Br. J. Haematol. 90:308-320(1995).
 RL [8]
 RP SEQUENCE OF 1421-1540 FROM N.A.
 RX MEDLINE=94020842; PubMed=8414518;
 RA Forster A., Rabbits T.H.;
 RA "A method for identifying genes within yeast artificial chromosomes:
 RA application to isolation of MLL fusion cDNAs from acute leukaemia
 RA translocations.";
 RT Oncogene 8:3157-3160(1993).
 RN [9]
 RP CHROMOSOMAL TRANSLOCATION WITH GAST.
 RX MEDLINE=20183971; PubMed=10706619;
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
 RA Wilson R.B., Jones D.H., Aditya K., Leonard D.G.B., Kushner B.H.,
 RA Williams T.M., Lange B.J., Felix C.A.;
 RA "Detection of leukemia-associated MLL-GAST translocation early during
 RA chemotherapy with DNA topoisomerase II inhibitors.";
 RT Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
 RL [10]
 RP CHROMOSOMAL TRANSLOCATION WITH AFBP21.
 RX MEDLINE=20115194; PubMed=10648423;
 RA Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;
 RA "Novel SH3 protein encoded by the AFBP21 gene is fused to the mixed
 RA lineage leukemia protein in a therapy-related leukemia with
 RA t(3;11)(p21;q23).";
 RT Blood 95:1066-1068(2000).
 RL [11]
 RP FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nucleus.
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT1/ENL;
 CC T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)
 CC THAT INVOLVES MLL AND MLLT3/AF3; T(6;11)(Q27;Q23) THAT INVOLVES
 CC MLL AND MLLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND
 CC MLLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLLT7/AFX1;
 CC T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLLT10/AF10;

CC T(1;11)(Q21;Q23) THAT INVOLVES MLL AND AFI10; T(11;19)(Q23;P13.3)
 CC that involves MLL and ELL; t(11;19)(q23;p23) that involves MLL
 CC and GAST; and t(3;11)(p21;q23) that involves MLL and AFBP21.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
 CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/genes/MLL.html";
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L04284; AAA58669.1; -;
 DR EMBL: 269744; CAA93625.1; -;
 DR EMBL: 269745; CAA93625.1; JOINED.
 DR EMBL: 269746; CAA93625.1; JOINED.
 DR EMBL: 269747; CAA93625.1; JOINED.
 DR EMBL: 269748; CAA93625.1; JOINED.
 DR EMBL: 269749; CAA93625.1; JOINED.
 DR EMBL: 269750; CAA93625.1; JOINED.
 DR EMBL: 269751; CAA93625.1; JOINED.
 DR EMBL: 269752; CAA93625.1; JOINED.
 DR EMBL: 269753; CAA93625.1; JOINED.
 DR EMBL: 269754; CAA93625.1; JOINED.
 DR EMBL: 269755; CAA93625.1; JOINED.
 DR EMBL: 269756; CAA93625.1; JOINED.
 DR EMBL: 269757; CAA93625.1; JOINED.
 DR EMBL: 269758; CAA93625.1; JOINED.
 DR EMBL: 269759; CAA93625.1; JOINED.
 DR EMBL: 269760; CAA93625.1; JOINED.
 DR EMBL: 269761; CAA93625.1; JOINED.
 DR EMBL: 269762; CAA93625.1; JOINED.
 DR EMBL: 269763; CAA93625.1; JOINED.
 DR EMBL: 269764; CAA93625.1; JOINED.
 DR EMBL: 269765; CAA93625.1; JOINED.
 DR EMBL: 269766; CAA93625.1; JOINED.
 DR EMBL: 269767; CAA93625.1; JOINED.
 DR EMBL: 269768; CAA93625.1; JOINED.
 DR EMBL: 269769; CAA93625.1; JOINED.
 DR EMBL: 269770; CAA93625.1; JOINED.
 DR EMBL: 269771; CAA93625.1; JOINED.
 DR EMBL: 269772; CAA93625.1; JOINED.
 DR EMBL: 269773; CAA93625.1; JOINED.
 DR EMBL: 269774; CAA93625.1; JOINED.
 DR EMBL: 269775; CAA93625.1; JOINED.
 DR EMBL: 269776; CAA93625.1; JOINED.
 DR EMBL: 269777; CAA93625.1; JOINED.
 DR EMBL: 269778; CAA93625.1; JOINED.
 DR EMBL: 269779; CAA93625.1; JOINED.
 DR EMBL: 269780; CAA93625.1; JOINED.
 DR EMBL: 269781; CAA93625.1; JOINED.
 DR EMBL: D14540; BAA03407.1; -;
 DR EMBL: L01986; AAA92511.1; -;
 DR EMBL: U04737; AAA18644.1; -;
 DR EMBL: S78570; AAA34770.1; -;
 DR EMBL: X83604; CAA58584.1; -;
 DR EMBL: S66432; AAB28545.1; -;
 DR EMBL: AFB23198; AAG26332.2; ALT_TERM.
 DR TRANSFAC: T02337; -;
 DR Genew: HGNC:7132; MLL.
 DR MIM: 159555; -;
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR003889; Fitch_C.
 DR InterPro: IPR003888; Fitch_N.
 DR InterPro: IPR003616; POSTSET.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR002857; Znf_CXXC.
 DR InterPro: IPR001965; Znf_PHD.


```

QY 30 ----- 29
Db 2489 DKGLSMGVKRAPMVGSAKKEIQAPRKRTVKTTLPLKMESESOKNAKLESSPASPL 2548
QY 30 ----- 29
Db 2549 QIESTSPTEPISASENPGDGPVAPSPNNNTSCDSSQNNQNLPLYODRNLMLDGPKEPQE 2608
QY 30 ----- 29
Db 2609 DGEFKRRYPKRSARARSNMFEGLPLLYGVRSYGEEDIPFYSSSTGKKRKSAGQVGA 2668
QY 30 ----- 29
Db 2669 DDLSTDEDLYYNFTRTYISSGGERLASHNLFREEQCDLPKISQLDGVDGDESPT 2728
QY 30 ----- 29
Db 2729 SVTATRKSSQIPKRNCKENGENTLKLDRPEDAGEKEHYTKSSVGHKNPKMCHSVSR 2788
QY 30 ----- 29
Db 2789 VKTGGODLEAQLSLESSRVRHTSPSDKNLLDVTYNTLLKSDSDNNNSDCGNILPSD 2848
QY 30 ----- 29
Db 2849 IMDFVLKNTSMQALGSPSSSSSELLNLGEGGLGDSNREKMGLEVESQQLPTEPVD 2908
QY 30 ----- 29
Db 2909 SSVSSISASEQFELPLPLPSDLVLTTRSPYPSQNPRLAVISDGEKRVITTEKSA 2968
QY 30 ----- 29
Db 2969 SSESDFALLSPGVDPPPEGHMTDPHFIOGHMDAHISSPCGSEVGHGNNODLTENSSPT 3028
QY 30 ----- 29
Db 3029 PGLQVPSPVPIQONQYVNSTDSBPQISOINAAVQTPPHLKPATEKLIYVNOQMOP 3088
QY 30 ----- 29
Db 3089 YVLOTLPNGVTKIQLTSSVSTPSVMEVNTSVLPGMGGLTLTGLNPSLPTSGSLPSS 3148
QY 30 ----- 29
Db 3149 ASKGLPMSHHQHLHSFPAAATQSSFPNNISNPSGILLGVQPPDQLYVSESSORTDLS 3208
QY 30 ----- 29
Db 3209 TTVATPSSGLKRRPISRLQTRKKKKLAPSSPTSNINAPSDVYSNMTLINFTPSQLPNHPSL 3268
QY 30 ----- 29
Db 3269 LDGLSINTSSHTVPIIKRKSIMYFEFAPLLPQSVGTATAGTSTISODTSHLS 3328
QY 30 ----- 29
Db 3329 GSVGLASSSVLVNVMQTTTPTTSASVPGHVTLLPNLLGTPDIGISMLLIKASQO 3388
QY 30 ----- 29
Db 3389 SLGIDQPVALLPSSSGMFPQLGTSQPTTAITAASTICVLPSTQTTGTTAASPSGEADE 3448
QY 30 ----- 29
Db 3449 HYOLOHVNOLLASKTGIHSSORDLDASGPOVSNFTQVDAPNSMGLFQONKALISAVQAS 3508
QY 30 ----- 31
Db 3509 PTPSGGSPSSSGQSRASPSVPGPTKPKTKRF 3543

```

```

RESULT 14
PGEM_HUMAN
ID PGEM_HUMAN STANDARD: PRT: 4393 AA.
AC P98160; Q16287;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112994; PubMed=1730768;
RA kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin, and Colon;
RX MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG/perlecan). A glyceric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 1018-1472 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
RN [4]
RP SEQUENCE OF 892-1398 FROM N.A.
RC TISSUE=Fibroblastoma;
RX MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1-3p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
RN [5]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=6234307;
RA Cohen I.R., Graesssel S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
CC -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.

```

CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X62515; CAA44373.1; -;
 DR EMBL; M85289; AAA52700.1; -;
 DR EMBL; M64283; AAA52699.1; -;
 DR EMBL; S76436; AAB21121.2; -;
 DR EMBL; L22078; -; NOT_ANNOTATED_CDS.
 DR HSP; P00740; 1EDM.
 DR Stena-2DPAGE; P98160; -;
 DR Genew; HGNC:5273; HSPG2.
 DR MIM; 142461; -;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF 2.
 DR InterPro; IPR001438; EGF II.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00047; Ig; 22.
 DR Pfam; PF00052; Laminin_B; 3.
 DR Pfam; PF00053; Laminin_EGF; 7.
 DR Pfam; PF00054; Laminin_G; 3.
 DR Pfam; PF00057; LDL_recept_a; 4.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00010; EGFBL00D.
 DR ProDom; PD003031; Laminin_B; 3.
 DR SMART; SM00180; EGF_Iam; 6.
 DR SMART; SM00408; EGF_Like; 8.
 DR SMART; SM00192; IgC2; 22.
 DR SMART; SM00182; LDca; 4.
 DR SMART; SM00281; Lamb; 3.
 DR SMART; SM00282; Lamb; 3.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 9.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS50025; IAM_G_DOMAIN; 3.
 DR PROSITE; PS01209; LDLRA_1; 4.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS50024; SEA; 1.
 DR Signal; Basement membrane; Proteoglycan: Repeat; Glycoprotein; Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain; Extracellular matrix; EGF-like domain.
 KW Extracellular matrix; EGF-like domain.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 4393 BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE PROTEIN.
 FT DOMAIN 80 194 SEA.
 FT DOMAIN 197 236 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 283 321 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 323 361 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 366 405 LDL-RECEPTOR CLASS A 4.
 FT DOMAIN 405 506 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 523 532 LAMININ EGF-LIKE 1 (N-TERMINAL).
 FT DOMAIN 533 732 LAMININ DOMAIN IV 1 (DOMAIN III A).
 FT DOMAIN 733 765 LAMININ EGF-LIKE 1 (C-TERMINAL).
 FT DOMAIN 766 815 LAMININ EGF-LIKE 2.

FT	DOMAIN	816	873	LAMININ EGF-LIKE 3. (INCOMPLETE).
FT	DOMAIN	881	925	LAMININ EGF-LIKE 4 (N-TERMINAL).
FT	DOMAIN	926	935	LAMININ EGF-LIKE 5 (N-TERMINAL).
FT	DOMAIN	936	1127	LAMININ DOMAIN IV 2 (DOMAIN III B).
FT	DOMAIN	1128	1160	LAMININ EGF-LIKE 6 (C-TERMINAL).
FT	DOMAIN	1161	1210	LAMININ EGF-LIKE 7.
FT	DOMAIN	1211	1267	LAMININ EGF-LIKE 8.
FT	DOMAIN	1277	1326	LAMININ EGF-LIKE 9 (N-TERMINAL).
FT	DOMAIN	1327	1336	LAMININ EGF-LIKE 10.
FT	DOMAIN	1337	1531	LAMININ EGF-LIKE 11.
FT	DOMAIN	1532	1564	LAMININ EGF-LIKE 12.
FT	DOMAIN	1565	1614	LAMININ EGF-LIKE 13.
FT	DOMAIN	1615	1672	LAMININ EGF-LIKE 14.
FT	DOMAIN	1679	1773	LAMININ EGF-LIKE 15.
FT	DOMAIN	1774	1867	LAMININ EGF-LIKE 16.
FT	DOMAIN	1868	1957	LAMININ EGF-LIKE 17.
FT	DOMAIN	1958	2053	LAMININ EGF-LIKE 18.
FT	DOMAIN	2054	2153	LAMININ EGF-LIKE 19.
FT	DOMAIN	2154	2246	LAMININ EGF-LIKE 20.
FT	DOMAIN	2247	2342	LAMININ EGF-LIKE 21.
FT	DOMAIN	2343	2438	LAMININ EGF-LIKE 22.
FT	DOMAIN	2439	2535	LAMININ EGF-LIKE 23.
FT	DOMAIN	2536	2631	LAMININ EGF-LIKE 24.
FT	DOMAIN	2632	2728	LAMININ EGF-LIKE 25.
FT	DOMAIN	2729	2828	LAMININ EGF-LIKE 26.
FT	DOMAIN	2829	2926	LAMININ EGF-LIKE 27.
FT	DOMAIN	2927	3023	LAMININ EGF-LIKE 28.
FT	DOMAIN	3024	3114	LAMININ EGF-LIKE 29.
FT	DOMAIN	3115	3213	LAMININ EGF-LIKE 30.
FT	DOMAIN	3214	3300	LAMININ EGF-LIKE 31.
FT	DOMAIN	3301	3401	LAMININ EGF-LIKE 32.
FT	DOMAIN	3402	3490	LAMININ EGF-LIKE 33.
FT	DOMAIN	3491	3576	LAMININ EGF-LIKE 34.
FT	DOMAIN	3577	3664	LAMININ EGF-LIKE 35.
FT	DOMAIN	3665	3845	LAMININ EGF-LIKE 36.
FT	DOMAIN	3846	3883	LAMININ EGF-LIKE 37.
FT	DOMAIN	3884	3924	LAMININ EGF-LIKE 38.
FT	DOMAIN	3925	4105	LAMININ EGF-LIKE 39.
FT	DOMAIN	4106	4143	LAMININ EGF-LIKE 40.
FT	DOMAIN	4144	4178	LAMININ EGF-LIKE 41.
FT	DOMAIN	4203	4391	LAMININ EGF-LIKE 42.
FT	DOMAIN	4392	4491	LAMININ EGF-LIKE 43.
FT	DOMAIN	4492	4591	LAMININ EGF-LIKE 44.
FT	DOMAIN	4592	4691	LAMININ EGF-LIKE 45.
FT	DOMAIN	4692	4791	LAMININ EGF-LIKE 46.
FT	DOMAIN	4792	4891	LAMININ EGF-LIKE 47.
FT	DOMAIN	4892	4991	LAMININ EGF-LIKE 48.
FT	DOMAIN	4992	5091	LAMININ EGF-LIKE 49.
FT	DOMAIN	5092	5191	LAMININ EGF-LIKE 50.
FT	DOMAIN	5192	5291	LAMININ EGF-LIKE 51.
FT	DOMAIN	5292	5391	LAMININ EGF-LIKE 52.
FT	DOMAIN	5392	5491	LAMININ EGF-LIKE 53.
FT	DOMAIN	5492	5591	LAMININ EGF-LIKE 54.
FT	DOMAIN	5592	5691	LAMININ EGF-LIKE 55.
FT	DOMAIN	5692	5791	LAMININ EGF-LIKE 56.
FT	DOMAIN	5792	5891	LAMININ EGF-LIKE 57.
FT	DOMAIN	5892	5991	LAMININ EGF-LIKE 58.
FT	DOMAIN	5992	6091	LAMININ EGF-LIKE 59.
FT	DOMAIN	6092	6191	LAMININ EGF-LIKE 60.
FT	DOMAIN	6192	6291	LAMININ EGF-LIKE 61.
FT	DOMAIN	6292	6391	LAMININ EGF-LIKE 62.
FT	DOMAIN	6392	6491	LAMININ EGF-LIKE 63.
FT	DOMAIN	6492	6591	LAMININ EGF-LIKE 64.
FT	DOMAIN	6592	6691	LAMININ EGF-LIKE 65.
FT	DOMAIN	6692	6791	LAMININ EGF-LIKE 66.
FT	DOMAIN	6792	6891	LAMININ EGF-LIKE 67.
FT	DOMAIN	6892	6991	LAMININ EGF-LIKE 68.
FT	DOMAIN	6992	7091	LAMININ EGF-LIKE 69.
FT	DOMAIN	7092	7191	LAMININ EGF-LIKE 70.
FT	DOMAIN	7192	7291	LAMININ EGF-LIKE 71.
FT	DOMAIN	7292	7391	LAMININ EGF-LIKE 72.
FT	DOMAIN	7392	7491	LAMININ EGF-LIKE 73.
FT	DOMAIN	7492	7591	LAMININ EGF-LIKE 74.
FT	DOMAIN	7592	7691	LAMININ EGF-LIKE 75.
FT	DOMAIN	7692	7791	LAMININ EGF-LIKE 76.
FT	DOMAIN	7792	7891	LAMININ EGF-LIKE 77.
FT	DOMAIN	7892	7991	LAMININ EGF-LIKE 78.
FT	DOMAIN	7992	8091	LAMININ EGF-LIKE 79.
FT	DOMAIN	8092	8191	LAMININ EGF-LIKE 80.
FT	DOMAIN	8192	8291	LAMININ EGF-LIKE 81.
FT	DOMAIN	8292	8391	LAMININ EGF-LIKE 82.
FT	DOMAIN	8392	8491	LAMININ EGF-LIKE 83.
FT	DOMAIN	8492	8591	LAMININ EGF-LIKE 84.
FT	DOMAIN	8592	8691	LAMININ EGF-LIKE 85.
FT	DOMAIN	8692	8791	LAMININ EGF-LIKE 86.
FT	DOMAIN	8792	8891	LAMININ EGF-LIKE 87.
FT	DOMAIN	8892	8991	LAMININ EGF-LIKE 88.
FT	DOMAIN	8992	9091	LAMININ EGF-LIKE 89.
FT	DOMAIN	9092	9191	LAMININ EGF-LIKE 90.
FT	DOMAIN	9192	9291	LAMININ EGF-LIKE 91.
FT	DOMAIN	9292	9391	LAMININ EGF-LIKE 92.
FT	DOMAIN	9392	9491	LAMININ EGF-LIKE 93.
FT	DOMAIN	9492	9591	LAMININ EGF-LIKE 94.
FT	DOMAIN	9592	9691	LAMININ EGF-LIKE 95.
FT	DOMAIN	9692	9791	LAMININ EGF-LIKE 96.
FT	DOMAIN	9792	9891	LAMININ EGF-LIKE 97.
FT	DOMAIN	9892	9991	LAMININ EGF-LIKE 98.
FT	DOMAIN	9992	10091	LAMININ EGF-LIKE 99.
FT	DOMAIN	10092	10191	LAMININ EGF-LIKE 100.

Query Match 38.5%; Score 60; DB 1; Length 4393;
 Best Local Similarity 0.5%; Pred. No. 1.8e+03;
 Matches 19; Conservative 3; Mismatches 9; Indels 3672; Gaps 5;

QY	1	SRXH	-----	4
DB	648	SRGHTPTPGALNQROVQSEHWHVHESGRPYQABELLQVLSLEAVLIQFYVNTKMASV	707	
QY	5	-----XHSME-----	9	
DB	708	GLSDIAMDYTVTHATSHGRAHVSERCPIGYSGLSCSCDAHFTVPGGPYLGTCSGCS	767	
QY	10	-----	9	
DB	768	CNGHASSCDPVYGHCLNQHNTBGPCKCKAGFFQDAKATATSCRPCTYDASRRF	827	

QY 10 ----- 9
Db 828 SPTCFIDTDGATCDACACAGYTGRCSCAPGEGNPIDPGKCAPVNOEIVRCDERGS 887
QY 10 ----- 9
Db 888 GTSSEACRCKNNVGRCLNECADRSFHLSTRNDGCLCFMGVSRHCTSSSWRAOLHG 947
QY 10 ----- 9
Db 948 ASEERHESLTMAASTHTNEGIFSP7PBELOFSSFHLLSGPYWSLPSRFLGDKVTSY 1007
QY 10 ----- 9
Db 1008 GGELEFVTOQSOPSTPLHGQPLVVLQGNITLLEHHVAQEPSGQSTFIVFREQAQ 1067
QY 10 ----- 9
Db 1068 RPDGQATREHLLMALAGIDTLLIRASYAQAPAESRVSISMDVAVPEBTGDPALVEBQ 1127
QY 10 ----- 9
Db 1128 CSCPGRGSPSCQDCTGTTRTPSGLYLGTCEKSCSHGSEACEPETGACQCGHHTBP 1187
QY 10 ----- 9
Db 1188 REQCPGYGDAQNGTPODCOLCPCYGDPAAQAHTCFDLDTHPTCDACSPGSHGRH 1247
QY 10 ----- 9
Db 1248 CERKAPGYGNPSQGPCORDSOVPEP1GICNCDPGSVSSQDAAQCCKAQBELTCS 1307
QY 10 ----- 9
Db 1308 HCRPHFHLASNPDCLEPCFCGIGITQOCASAYTRHLSTHAPDQGFALVNPQRNS 1367
QY 10 ----- 9
Db 1368 RLTEFTVEPVEGAQLSEGNFAQLGHSFYQLPETYQDRAVAAVYGLRYTLSTYAGP 1427
QY 10 ----- 9
Db 1428 QGSPLSDPDVQITGNINMLVASOPALQGPERSYEIMREERWRPDGQATREHLLMAL 1487
QY 10 ----- 9
Db 1488 ADDELIRATSSVPLVASISAVSLEVAQGPSPNRPALVEBECKPCPGYIGLSQDCA 1547
QY 10 ----- 9
Db 1548 PGYTRTSGLYLGHCELCNCHSHDLCHPBTGACSOQHNAAGEFCELCAPGYGDATAG 1607
QY 10 ----- 9
Db 1608 TPEDQPCACPLTNPENMSRICEISLACGYRCTACEPGTQCYCEQCPGYVGNPSVQ 1667
QY 10 ----- 9
Db 1668 GQCLPFTNOAPLVEVEYHPARSIVPGGSHSLRCQVSGRPHFYFWSREDGRVPSTQOR 1727
QY 10 ----- 9
Db 1728 HOGSELHFPSPVPSDAGVYICTCRNLHRSNTSRAELLVTEAPSKPITVVEQRSQVAP 1787
QY 10 ----- 9
Db 1788 GADVTICTAKSKSPAYTTLVWTRLANGLPTRAMDFNGILITIRNOLSDAGTYVCTGSNM 1847
QY 10 ----- 9
Db 1848 FAMDQGTATLHQASGSLASAPVYSIHPQULVQGLAEFRCSATGSPPTLEWTGGPGG 1907
QY 10 ----- 9

Db 1908 QLPAAQIHGILRLPAVEPTDQAOYLCAHSSAGQAVARAVLHVGGGPRVQVSPERT 1967
QY 10 ----- 9
Db 1968 QVHAGRTVRLYCRAGVPSATITWRKEGSLPPOANSEPTDIATLIPAITTADAGHYLC 2027
QY 10 ----- 9
Db 2028 VATSPAGTAQARIQVVLASASDAQPPVKIESSPSVTEGOTDLNLCVAGSAHAQVWY 2087
QY 10 ----- 18
Db 2088 RRGSLPHHTQVHGSRLRLPDQVSPADSGEYCVRENGSGPEASTIVSVLHGTGSPSYT 2147
QY 19 ----- 18
Db 2148 FVPGSTRPIRIEPSSSHVABGOTDLNLCVPGQAAQVTHMKRGSLPARHQTHGSLRL 2207
QY 19 ----- 18
Db 2208 HQVTPADSGEYVCHVGTSGPLEASVLTIEASYIPGPIPPRIESSSTVABGOTDLNLS 2267
QY 19 ----- 18
Db 2268 CVVAGAAQVTHYKRGSLPARHQVHGSRLYIFQASPADAGQVCHASNGMEASTIVTV 2327
QY 19 ----- 18
Db 2328 TGTQGANIATYAGSTQPIRIEPPSSQVABGOTDLNLCVPGQAAQVTHMKRGSLPYRH 2387
QY 19 ----- 18
Db 2388 QTHGSLRLYQASPADSGEYCVILGSSVPLEASVLTIEPAGSPALGVTPIVRIESS 2447
QY 19 ----- 18
Db 2448 SOVAGOTDLNLCVAGAAQVTHMKRGSLPARHQVHGSRLQLQVTPADSGEYCVRY 2507
QY 19 ----- 18
Db 2508 VGSSTQASVLYTIQRLSGSHSQVAYPIVRISSASLANGHTLNLCLVASQAPHTI 2567
QY 19 ----- 18
Db 2568 TWYKRGSLPSRHQIVGSLRLIPQVTPADSGEYCHVSNAGSRETSLIYIQSGSSHV 2627
QY 19 ----- 18
Db 2628 PRVSPPIRIESSPTVVGOTDLNLCVVARQPAITWYKRGSLPSRHQTHGSLRLHQ 2687
QY 19 ----- 18
Db 2688 MSVADSGEYVCRRANNIDALASIVISVPSAGSPSAPGSSMPIRISSSHVABGETLD 2747
QY 19 ----- 18
Db 2748 LNCVVGQAHAQVTHMKRGSLPSYHQTGSLRLHLHVSPADSGEYCVRWGSSGLEAS 2807
QY 19 ----- 18
Db 2808 VLVTEASGSSAVHPADGAPPIRIEPPSSRAVEGOTDLKCVVPGQAHAQVTHMKRG 2867
QY 19 ----- 18
Db 2868 NLPARHQVHGPLRLNOVSPADSGEYSCQVYTGSSGLEASVLTIESSSPQIPAPQLAQ 2927
QY 19 ----- 20
Db 2928 PIYIASSSHVTEGOTDLNLCVPGQAHAQVTHYKRGSLPARHQTHGSLRLHHSVPAD 2987
QY 21 ----- 20

```

Db 2968 GGEVYCRAGGPGPRQASFTVYVPPSSGSSYRLSPVTSIDPPSTVOGQODAFKCL 3047
QY 21 ----- 20
Db 3048 HDGAAPISLEKTRNOELEDNVHISPNSSITTYGTRPSNHGTRYCAVSAANVGAQSVN 3107
QY 21 ----- 20
Db 3108 LSVHGPTVSVLPBGPVWVKVAKVTLLECYSAGBRSARMTIRISIPAKLEQRTYGLMD 3167
QY 21 ----- 20
Db 3168 SHTVLQISSANPSDAGTYVCLAAQNALGTAQKOVEIVDTGAMADGADQVAAEAEALYVEA 3227
QY 21 ----- 20
Db 3228 GHTATLRCSATGSPARTIHMSKLSPLFMQHRLEGDTLIPRAVAQDSGQYTCNATSPAG 3287
QY 21 ----- 20
Db 3288 HAEATIIILHVESPPYATTVPEHASVQAGETVQLOCLAHGTPPLTFQMSRVGSSLPRGATA 3347
QY 21 ----- 20
Db 3348 RNEILHFERAAPEDSGRYRCRYTKVGSAAEAPQALLVOGPPGSLPATISIPAGSTPTVOVT 3407
QY 21 ----- 20
Db 3408 POLETKSIGASVEFHCAVPSDRGTQLRMFKREGQLPPGHVSQDVLRINQNLDSQCGTYI 3467
QY 21 ----- 20
Db 3468 QQAHGPMCKQAQSAQVLYQALPSVLIINIRTSVQTVVGVHAEFECLALGDPKQVTVSKV 3527
QY 21 ----- 20
Db 3528 GGHLPRTIGVSGVYVRIAHVLAADAGQRCYATNAAGTQSHVLLVQALPOLSMPOEVR 3587
QY 21 ----- 20
Db 3588 VPAGSAVFPCLASGYPTPDISMKLDGSLPPDSRLNNMLPSVQPDAGTYVCTATN 3647
QY 21 ----- 20
Db 3648 ROGKYKAPAHLOVERVVPYFTQTPYSFLPLPTIKAYRKEIKITERPDSADGMLLYNG 3707
QY 21 ----- 20
Db 3708 QKRVPGSPTNLANRQDPDISFGLVGGRPERFDAGSGMATIRHPTPLALGHFHTVTLRS 3767
QY 21 ----- 20
Db 3768 LTOGLIYVDLAPVNGTSQGFQGLDINELYLIGPYDGAIPKAGLSGFGICVRELRI 3827
QY 21 ----- 20
Db 3828 QGEIVTHDNLTLAHGISHCPTCDRPCQNGCHDESSSYVCVCPAGFTGSRCEHSQA 3887
QY 21 ----- 20
Db 3888 LHCPEACGPDATCVNRPDGRGTCRCHLGRSGLRCEGVTYVTPSLSGAGSYIALPALT 3947
QY 21 ----- 20
Db 3948 NTHHELRLDVEFKPLADGVLLFSGKSGPEDEVSILANVGHLEFRYELSGIAYLRTA 4007
QY 21 ----- 28
Db 4008 EPLALGMRHRYSAERLNKDSLRVNGRPLVRSSPGKSGGLNHTLLYLGGVPSVPLSP 4067
QY 29 ----- 28
Db 4068 ATNMSAHPRGCVGVNKGRLDLYSFLSQSGIGQCYDSSPCEROPCQHATCMPAGEY 4127

```

```

QY 29 ----- 28
Db 4128 EFQCLCRDGIKGDLCHEENPCQLREPCILHGTCQGTCLCLPSPGRCQGGSGHIAE 4187
QY 29 ----- 28
Db 4188 SDWHLGSGGNDAPGOYGAFFHDDGLAFPGHVSLSLPEVETIELEVRTSTASGLILM 4247
QY 29 ----- 28
Db 4248 QGVEVEAGQGDPLISLQDGHLYFRYQLGSGEARLYSEDEINDGEMHRYTALREGRG 4307
QY 29 ----- 31
Db 4308 SIQVDEELVSGRSRPGPNVNAKSGITYIGAPDVATLTGGRF 4350

RESULT 15
DYHB.CHLRE
ID DYHB.CHLRE STRAND: PRT: 4568 AA.
AC 039565:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE Dynein beta chain, flagellar outer arm.
GN ODA4 OR ODA-4 OR SUP1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21gr;
RX MEDLINE=94274778; PubMed=8006077;
RA Mitchell D.R.; Brown K.S.;
RT "Sequence analysis of the Chlamydomonas alpha and beta dynein heavy
   chain genes."
RL J. Cell Sci. 107:635-644(1994).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
   FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYNEIN HAS ATPASE ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND
   GAMMA), 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: U02963; AAA19956.1;
DR InterPro: IPR004273; Dynein_heavy.
DR Pfam: PF03028; Dynein_heavy; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
   Coiled coil.
FT FT DOMAIN 277 293 COILED COIL (POTENTIAL).
FT FT DOMAIN 1158 1175 COILED COIL (POTENTIAL).
FT FT DOMAIN 1372 1400 COILED COIL (POTENTIAL).
FT FT DOMAIN 1614 1650 COILED COIL (POTENTIAL).
FT FT DOMAIN 1778 1825 COILED COIL (POTENTIAL).
FT FT DOMAIN 2017 2045 MICROTUBULE-BINDING (POTENTIAL).
FT FT DOMAIN 2831 2848 COILED COIL (POTENTIAL).
FT FT DOMAIN 3106 3162 COILED COIL (POTENTIAL).
FT FT DOMAIN 3339 3425 COILED COIL (POTENTIAL).
FT FT DOMAIN 3648 3728 COILED COIL (POTENTIAL).
FT FT NP_BIND 1919 1926 ATP (POTENTIAL).
FT FT NP_BIND 2202 2209 ATP (POTENTIAL).
FT FT NP_BIND 2530 2537 ATP (POTENTIAL).
FT FT NP_BIND 2879 2886 ATP (POTENTIAL).

```


SQ SEQUENCE 4568 AA; 519961 MW; 9A9A5393C7C36AE7 CRC64;
 Query Match 38.5%; Score 60; DB 1; Length 4568;
 Best Local Similarity 0.8%; Pred. No. 2e+03;
 Matches 19; Conservative 3; Mismatches 6; Indels 2209; Gaps 6;
 QY 4 HXHSNE----- 9
 Db 1704 HHSKEGEVEFEEDSCDGPVEVWLVQNVDSMKLALQVEFRKALPTVDLPRTKWIYV 1763
 QY 10 -----XRT----- 12
 Db 1764 SNAQTVVSRPTFTQIEINAEADDEEGNEBALKELDROVOQLADLIDEINKQTSIDRK 1823
 QY 13 ----- 12
 Db 1824 KLITCTIDVSRDLVQKLIDERVEQMCFOQOSQRLRIQSEKTKTCQVNICDAEIAVSY 1883
 QY 13 ----- 12
 Db 1884 EYICNGCGLCTPTLDRCFITLQAQRLVIGAPAGPAGTGKTETTKDLARALGIQCYVF 1943
 QY 13 ----- 12
 Db 1944 NCSQDMYKAMGHYKGLAQGTAMCCEDEFNRIPYAVLSVSTQYKTVLDIRAKKERFT 2003
 QY 13 ----- 12
 Db 2004 FEDADISLSTVMAFITNPGYPGRALPESLKALFRPVSMVPDLALICEIMLMAGFQ 2063
 QY 13 ----- 12
 Db 2064 MSKILSRKFLVILKCEBLSKSRHYDKLRAITTLVAGMKRAAPSEDRVLLRAL 2123
 QY 13 -----PD----- 14
 Db 2124 RDNIGKLTDADTSIFMGLNDLPKTELEVPRALDKAFDEAKKAPTELGYQDDDFL 2183
 QY 15 -----INP----- 17
 Db 2184 KISHVRELTVYRMSVFLGAGCGKTAVMWTLTRAQNSGSEKITTYQAVNPRAVTRNELYG 2243
 QY 18 ----- 17
 Db 2244 YLHPATREMGKLSVTFERNANKNKHQIYLDGDIADMEIESMNTVADNKMILLAS 2303
 QY 18 ----- 17
 Db 2304 NERIPPSMRLLLEINHMVHCSPATVSRGVIFINADVQMPVVASWIDKLEAEYRP 2363
 QY 18 ----- 17
 Db 2364 LITLALFKYVDPCLCHGRNFKYVPLPAVNOAMTICKILEGILPKETVGAPPPDKLL 2423
 QY 18 ----- 17
 Db 2424 HYHVFACVMAFGCMLVDKVDYRTQFSKMWVSEMKDQFPEKGLVYDYVDQNCIMY 2483
 QY 18 ----- 17
 Db 2484 PWEDRVTKFOYIPEDPTSLFPVETVTRLTFLDLSLVSNKHVAMFVGNMGTGSAIMVNK 2543
 QY 18 ----- 17
 Db 2544 LRNMDETMSFYTIMNSLSEAPALQVILEQPLEKSGVYRPGSRMYFVDDANMPL 2603
 QY 18 -----AMT----- 20
 Db 2604 VDKYDTOSIELLRQWVDYHGWDKVKIQLKEIINCQMACMNPAGSFNITPRQRRHV 2663
 QY 21 ----- 20
 Db 2664 TFAVQMPNAETIRAMYQIITDGHFSFDDVVAKMSNKLVDATCELHRVNMHFLPSAVKR 2723

QY 21 ----- 20
 Db 2724 HYQNLRLDSNITQGLTRAIKEYYREFVKVARIWVEHCERVFDRMINEADMAKPFDEFY 2783
 QY 21 ----- 20
 Db 2784 AVTKKFFDDCGMAVIERPLIYASHASMTYTEDBVPVYALSSYDLRKTELDKREYN 2843
 QY 21 ----- 20
 Db 2844 ESNAVMDLVLFQAMEHVTBARIIDLPRGNAMLVGSGSKSLARLASIYIGYEYQI 2903
 QY 21 ----- 20
 Db 2904 SVSSTYINDKENLGLYRKAGTKGTPITFLMTDQIVKEGLYINDLSTGYADLF 2963
 QY 21 ----- 20
 Db 2964 TPEDKEAFNAVRNEVKAAGILDSANCMDFIDKYRKFLHYLCSPVGDKFRIRAROF 3023
 QY 21 ----- 20
 Db 3024 PALVNCMTFDMFGWGEALVSAQRFVDPNMEVVENIAYHMAVAHOCVSEASERF 3083
 QY 21 ----- 20
 Db 3084 KEAFRRNTTPKSYELISLYKMLQLKRDILRSKERLENGIDKIAQAAQVTDLQRY 3143
 QY 21 ----- 20
 Db 3144 LKEEQIVDEKKAQDELIVSIGEKAIVDOAVEAGREDEEATALQTEVSAFOACEKD 3203
 QY 21 ----- 20
 Db 3204 LLEAEPLIAQEAALNSLKKELSELKSFSPAELIYVAAACLVLCGKIPKDRDWA 3263
 QY 21 ----- 20
 Db 3264 GKEMADVNSFLSLNPFDDKDVYVVEVEKDYISNPGFTPDNIKKSACAGLCSWV 3323
 QY 21 ----- 20
 Db 3324 INICKYFRIVYVAPRRALAEANKLIDPANKKLVIRDEVRLQDRVALBQSIMKATE 3383
 QY 21 ----- 20
 Db 3384 DKMAITAQADRTRARKQMAERLINGSENTRMGAELKRLSELEGRVGDVLIASAFVSY 3443
 QY 21 -----XXRGIRP----- 27
 Db 3444 AGPFNNQFRKSLVEKWLPIIERQIPMTQGIKRPDLDLTDATKAKMANEGLPDPLSV 3503
 QY 28 ----- 27
 Db 3504 NGALMSNASRMALMDPOLQIKWIIINFTNGLVITQOSOPKYIDQYINCIENGWPLLI 3563
 QY 28 ----- 27
 Db 3564 ENLPVIDAVLDPVIGKMTIKKGRNIMKIGDAEVOYDSRFLYIQTLSNPHKPEVAA 3623
 QY 28 ----- 27
 Db 3624 QFTLVNFCYTERGLEDDLALVVDHERPDLOEAGLVSLNEYNITLVELENNLFFVLA 3683
 QY 28 ----- 27
 Db 3684 NATGNILENIELLEGLBETKRTAVELIEEKVLAKQTEIQAIAKAREVYPVATRSGLYFL 3743
 QY 28 ----- 27
 Db 3744 IDNLALDRVYHYMANEFVFLKKGMDMTPGGKDESKVPLAERLQNEVDLDRKVELLEVET 3803

QY	28	-----	27
Db	3804	TCFVLIGVAGLFRHRKLIYAFQLCMOILRSRGELHYAKFEYILLRGPKVMGADNPLHDW	3863
QY	28	-----	27
Db	3864	VSDSVWGSVQALKELEDYQGLPEDLIGSSKRWREMELEPDEPLPGDWKRMOEFDKL	3923
QY	28	-----YGRF	31
Db	3924	LFRALRPDRLTSAMGRF	3940

Search completed: May 1, 2003, 14:47:12
 job time : 18 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 10.1719 Seconds

(without alignments)
126,404 Million cell updates/sec

Title: US-09-446-543a-61

Perfect score: 171

Sequence: 1 SRTHRSMEIRPDINPAMWASRGIRPYGRF 31

Scoring table: BLOSUM62

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	171	100.0	87	PRRP_HUMAN	P81277 homo sapien
2	158	92.4	98	PRRP_BOVIN	P81264 bos taurus
3	149	87.1	83	PRRP_RAT	P81278 rattus norv
4	54	31.6	428	NER3_HUMAN	Q9uq49 homo sapien
5	52	30.4	288	Y587_PASMO	Q9cns56 pasteurilla
6	50.5	29.5	1882	Y468_MYCPN	Q97859 mycoplasma
7	50	29.2	428	NER3_BOVIN	Q97859 bos taurus
8	47.5	27.8	345	NER3_BOVIN	Q97859 bos taurus
9	47.5	27.8	533	CNC_DROME	P20482 drosophila
10	47	27.5	402	EX7L_STRCO	Q97853 streptomyce
11	47	27.5	676	EX7L_HUMAN	Q92935 homo sapien
12	46.5	27.2	240	PLSC_HELPY	Q25903 heliobacte
13	46.5	27.2	1693	POLN_HEVBU	P29324 hepatitis e
14	46.5	27.2	1693	POLN_HEVBU	Q04610 hepatitis e
15	46.5	27.2	1783	Y468_MYCCE	Q49460 mycoplasma
16	46	26.9	342	Y762_METUA	Q58172 methanococc
17	46	26.9	347	Y576_METUA	Q57996 methanococc
18	46	26.9	383	CYCR_CHRYI	Q82947 chromatiu
19	46	26.9	1400	RIF1_SCHPO	Q96up3 schizosacch
20	45.5	26.6	239	6PGL_SYNY3	P72052 mycobacteri
21	45.5	26.6	407	Y116_MYCVU	Q95176 saccharomyc
22	45	26.3	213	SRN2_YEAST	P10181 drosophila
23	45	26.3	350	HMNO_DROME	P29013 escherichia
24	45	26.3	510	YCGB_ECOLI	P18395 rattus norv
25	45	26.3	798	UNR_RAT	Q95443 methanobact
26	45	26.3	940	UVRA_VIBCH	Q26543 rhizodolum i
27	45	26.3	962	UVRA_METTH	Q26543 rhizodolum i
28	45	26.3	973	UVRA_RHILIO	Q26543 rhizodolum i
29	45	26.3	973	UVRA_RHIME	Q26543 rhizodolum i
30	44.5	26.0	860	VG12_BPBO3	P37893 bacterioph
31	44	25.7	137	SMR2_RAT	P18897 rattus norv
32	44	25.7	364	YAIW_ECOLI	P77562 escherichia
33	44	25.7	386	CRTY_AGRAU	P54974 agrobacteri

SEA ID NO: 61
Database: SwissProt-40
AC NO: P81277

34	44	25.7	476	1	YAU_ECOLI	P30143 escherichia
35	44	25.7	581	1	POL_MLYAK	P31795 radiation m
36	44	25.7	591	1	PYRG_HUMAN	P17812 homo sapien
37	44	25.7	719	1	NRPL_YEAST	P32770 saccharomyc
38	44	25.7	843	1	POL_MLYAK	P03357 akt murine
39	44	25.7	1087	1	POL_HUMAN	Q9u1a9 homo sapien
40	44	25.7	1087	1	RP16_MOUSE	Q9epk7 mus musculu
41	44	25.7	1196	1	POL_MLYAK	P03356 akt murine
42	44	25.7	1196	1	POL_MLYAK	P11227 radiation m
43	43.5	25.4	568	1	RSL_RICPR	Q9zd78 rickettsia
44	43.5	25.4	572	1	GAG_IPHA	P04023 hamster int
45	43.5	25.4	770	1	AVP3_ARATH	P31414 arabidopsis

ALIGNMENTS

RESULT 1						
ID	PRRP_HUMAN	STANDARD;	PRT;	87	AA.	
AC	P81277;					
DT	30-MAY-2000 (Rel. 39, Created)					
DT	30-MAY-2000 (Rel. 39, Last sequence update)					
DT	15-JUN-2002 (Rel. 41, Last annotation update)					
DE	Prolactin-releasing peptide precursor (PRP) (Prolactin-releasing hormone) [Contains: Prolactin-releasing peptide PRP31; Prolactin-releasing peptide PRP20].					
DE	PRP.					
GN	Homo sapiens (Human).					
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OX	NCBI_TaxID:9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE-Brain:					
RX	MEDLINE-98266781; Pubmed-9607765;					
RA	Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kikada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;					
RT	"A prolactin-releasing peptide in the brain."					
RL	Nature 393:272-276(1998).					
RN	[2]					
RP	TISSUE SPECIFICITY.					
RX	MEDLINE-99426652; Pubmed-10498338;					
RA	Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kikada C., Kurokawa T., Nishimura O., Onda H., Fujino M.;					
RT	"Tissue distribution of prolactin-releasing peptide (PRP) and its receptor."					
RL	Regul. Pept. 83:1-10(1999).					
CC	-1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.					
CC	-1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.					
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).					
DR	EMBL; AB015419; BAA29027.1;					
DR	MM: 602663;					
KW	Hormone; Amidation; Signal.					
FT	SIGNAL					
FT	PEPTIDE					
FT	MOD_RES					
FT	SEQUENCE					
Query Match	87	AA;	9639	MM;	229A2F350CF981B	CR664;
	100.0%;	Score 171;	DB 1;	Length 87;		

Best Local Similarity 100.0%; Pred. No. 7e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTSRHSMETRPDPINPAMVYASRGIRPVGRF 31
DB 23 SRTSRHSMETRPDPINPAMVYASRGIRPVGRF 53

RESULT 2

PRRP_BOVIN STANDARD; PRT; 98 AA.
AC P81264;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PrRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PrRP31; Proactin-releasing peptide PrRP20].
GN PRH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 23-52.
RT TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;

RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A proactin-releasing peptide in the brain."
RL Nature 393:272-276(1998).
CC -1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.
CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB015417; BAA29025.1;
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 22
FT PEPTIDE 23 53 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT PEPTIDE 33 53 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 53 53 AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ SEQUENCE 98 AA; 10544 MW; 08AC35A13B0FA908 CRC64;

Query Match 92.4%; Score 158; DB 1; Length 98;
Best Local Similarity 90.3%; Pred. No. 7.1e-17;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTSRHSMETRPDPINPAMVYASRGIRPVGRF 31
DB 23 SRTSRHSMETRPDPINPAMVYASRGIRPVGRF 53

RESULT 3

PRRP_RAT STANDARD; PRT; 83 AA.
AC P81276;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proactin-releasing peptide precursor (PrRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PrRP31; Proactin-releasing peptide PrRP20].

GN PRH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

RA Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A proactin-releasing peptide in the brain."
RL Nature 393:272-276(1998).
CC -----
DR EMBL; AB015418; BAA29026.1;
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 21
FT PEPTIDE 22 52 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT PEPTIDE 33 52 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 52 52 AMIDATION (G-53 PROVIDE AMIDE GROUP).
SQ SEQUENCE 83 AA; 9215 MW; DDC75A264EBE4F29 CRC64;

Query Match 87.1%; Score 149; DB 1; Length 83;
Best Local Similarity 83.9%; Pred. No. 1.3e-15;
Matches 26; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRTSRHSMETRPDPINPAMVYASRGIRPVGRF 31
DB 22 SRTSRHSMETRPDPINPAMVYASRGIRPVGRF 52

RESULT 4
NER3_HUMAN STANDARD; PRT; 428 AA.
AC Q9UQ49; Q9NOE1;
ID NER3_HUMAN
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sialidase 3 (EC 3.2.1.18) (Membrane sialidase) (Ganglioside sialidase) (N-acetyl-alpha-neuraminidase 3).
GN NEU3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=Brain;
RX MEDLINE=9935353; PubMed=10405317;
RA Wada T., Yoshikawa Y., Tokuyama S., Kuwabara M., Akita H., Miyagi T.;

RT Cloning, expression, and chromosomal mapping of a human ganglioside
 RT sialidase.
 RL Biochem. Biophys. Res. Commun. 261:21-27(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Skeletal muscle;
 RX PubMed-10861246;
 RA Monti E., Bassi M.T., Papini N., Riboni M., Manzoni M., Venerando B.,
 RA Croci G., Preti A., Ballabio A., Tettamanti G., Borsani G.;
 RT Identification and expression of NEU3, a novel human sialidase
 RT associated to the plasma membrane."
 RL Biochem. J. 349:343-351(2000).
 CC -1- FUNCTION: Plays a role in modulating the ganglioside content of
 CC the lipid bilayer at the level of membrane-bound sialyl
 CC glycoconjugates.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
 CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
 CC oligosaccharides; glycoproteins, glycolipids, colominic acid and
 CC synthetic substrates.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.
 CC -1- TISSUE SPECIFICITY: Highly expressed in skeletal muscle, testis,
 CC adrenal gland and thymus, followed by pancreas, liver, heart and
 CC thymus. Weakly expressed in kidney, placenta, brain and lung.
 CC -1- MISCELLANEOUS: Optimum pH is 3.8.
 CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB008185; BAA82611.1;
 DR EMBL: Y18563; CAB96131.1; ALT_INIT.
 DR Genew; HGNC:7760; NEU3.
 DR MIM: 604617;
 DR InterPro: IPR002860; GH_BNR.
 DR Pfam: PF02012; BNR. 6.
 KW Hydrolyase; Glycosidase; Membrane; Repeat.
 FT REPEAT 129 140
 FT REPEAT 203 214 BNR 1.
 FT REPEAT 254 265 BNR 2.
 FT REPEAT 27 27 FRIP MOTIF.
 FT SITE 24 25 BY SIMILARITY.
 FT ACT_SITE 25 25 POTENTIAL.
 FT ACT_SITE 45 45 POTENTIAL.
 FT ACT_SITE 50 50 POTENTIAL.
 FT ACT_SITE 87 87 POTENTIAL.
 FT ACT_SITE 225 225 POTENTIAL.
 FT ACT_SITE 245 245 POTENTIAL.
 FT ACT_SITE 340 340 BY SIMILARITY.
 FT ACT_SITE 370 370 POTENTIAL.
 FT ACT_SITE 387 387 POTENTIAL.
 SQ SEQUENCE 428 AA; 48252 MW; 3501DD9359A78C98 CRC64;
 QY 2 RTRHSMETRTDIPNPAWASRGIRPV 28
 DB 195 KTRPHSIMITYSDLGVTWHGRILRPV 221
 Query Match 31.6%; Score 54; DB 1; Length 428;
 Best Local Similarity 37.0%; Pred. No. 1.5;
 Matches 10; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

RESULT 5
 Y587_PASMU STANDARD; PRT; 288 AA.
 AC Q9CN56;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein PM0587.
 GN PM0587.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Pasteurella.
 OK NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Pm70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -1- SIMILARITY: BELONGS TO THE FRUCTOSAMINE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE006094; AAK02671.1;
 DR Hypothetical protein; Transferase; Kinase; Complete proteome.
 KW SEQUENCE 288 AA; 33778 MW; FAD2F6C26014D940 CRC64;
 SO
 QY 5 RSHSMETRTDIPNPAWASRGIRPV 28
 DB 20 KHKRIHTGHEHMAIIDDGIQPV 43
 Query Match 30.4%; Score 52; DB 1; Length 288;
 Best Local Similarity 37.5%; Pred. No. 1.9;
 Matches 9; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

RESULT 6
 Y468_MYCPN STANDARD; PRT; 1882 AA.
 AC P75109; O50317;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein MG468 homolog (K05_orf1882).
 GN MPN684 OR MP158.
 OS Mycoplasma pneumoniae.
 CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OK NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae."
 RL Nucleic Acids Res. 24:4420-4449(1996).
 RN [2]
 RP SEQUENCE OF 1-1848 FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=96177562; PubMed=8604303;
 RA Hilbert H., Himmelreich R., Plagens H., Herrmann R.;
 RT "Sequence analysis of 56 kb from the genome of the bacterium
 RT Mycoplasma pneumoniae comprising the dna region, the atp operon and a
 RT cluster of ribosomal protein genes."
 RL Nucleic Acids Res. 24:628-639(1996).
 RN [3]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=21088919; PubMed=11271496;
 RA Regula J.T., Ueberle B., Boguth G., Goery A., Schnoelzer M.,
 RA Herrmann R., Frank R.;
 RT "Towards a two-dimensional proteome map of Mycoplasma pneumoniae.";

```
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/termsand
CC or send an email to license@ebi.ac.uk).
```

```
DR EMBL; AB008184; BAA75071.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR_3.
KW Hydroxylase; Glycosidase; Membrane; Repeat.
FT REPEAT 129 140 BNR 1.
FT REPEAT 203 214 BNR 2.
FT REPEAT 254 265 BNR 3.
FT SITE 24 27 FRIP MOTIF.
FT ACT_SITE 25 25 BY SIMILARITY.
FT ACT_SITE 45 45 POTENTIAL.
FT ACT_SITE 50 50 POTENTIAL.
FT ACT_SITE 87 87 POTENTIAL.
FT ACT_SITE 225 225 POTENTIAL.
FT ACT_SITE 245 245 POTENTIAL.
FT ACT_SITE 341 341 BY SIMILARITY.
FT ACT_SITE 371 371 POTENTIAL.
FT ACT_SITE 388 388 POTENTIAL.
SQ SEQUENCE 428 AA; 47916 MW; 41B34FF3245A8F21 CRC64;
```

```
OY 2 RTHRSMEIRTPDINPARYASGRIPV 28
Db 195 RARPSLMIVSDGLGATWNGRLKPM 221
```

```
RESULT 8
ARGC_BACHD ID ARGC_BACHD STANDARD; PRI; 345 AA.
```

```
AC O9KR6V2;
AD 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGRP) (N-
DE acetyl-glutamate semialdehyde dehydrogenase) (MAGSA dehydrogenase).
GN AROC OR BH2900.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_taxID=86665; [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -I- CATALYTIC ACTIVITY: N-acetyl-L-glutamyl 5-semialdehyde + NADPH(+)
CC + phosphate = N-acetyl-L-glutamyl phosphate + NADPH.
CC -I- PATHWAY: Arginine biosynthesis; third step.
CC -I- SIMILARITY: BELONGS TO THE MAGSA DEHYDROGENASE FAMILY.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/termsand
CC or send an email to license@ebi.ac.uk).
```

```
EMBL; AP001517; BAB0619.1; -
DR InterPro; IPR000706; AGRP_act_site.
DR InterPro; IPR000534; Semialdh_ch.
```

```
QY 17 PAWYASRG-----IRPVG 29
      | | | | |
Db 89 PEWYAPRGQLSLRAAEIKPVG 109
```


QY 1 SRTHRSMEIRTPDINPAMY 20
DB 197 ARTRVMEISYTPDFNSPTWY 217

RESULT 13
POLN_HEVBU STANDARD: PRT: 1693 AA.

AC P29324;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Non-structural polypeptide [Contains: RNA-directed RNA polymerase
(EC 2.7.7.48); Helicase].
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=31767;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA Fry K.E., Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RT Virology 185:120-131(1991).
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M73218; AAA5734.1; -
CC PIR: A40778; MNMWH.
CC DR MEROPS: C41.001; -
CC DR InterPro: IPR002589; A1pp.
CC DR InterPro: IPR001230; Prenyl_site.
CC DR InterPro: IPR001788; RNA_dep_RNAPol2.
CC DR InterPro: IPR002588; V_methyltransf.
CC DR InterPro: IPR000606; Viral_helicase1.
CC DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
CC DR Pfam: PF01443; Viral_helicase1; 1.
CC DR Pfam: PF01660; Vmethyltransf; 1.
CC DR Pfam: PF01661; A1pp; 1.
CC DR SMART: SM00506; A1pp; 1.
CC KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
CC ATP-binding.
CC FT NP_BIND 975 982 ATP (POTENTIAL).
CC SEQUENCE 1693 AA; 185191 MW; 2F355E46E9ED219B CRC64;

Query Match 27.2%; Score 46.5; DB 1; Length 1693;
Best Local Similarity 37.0%; Pred. No. 89;
Matches 10; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

QY 2 RTHRSMEIRTPDINPAMY-ASRGIRP 27
DB 904 RNHRGDELTYLPFLAARFEANRPTRP 930

RESULT 14
POLN_HEVBU STANDARD: PRT: 1693 AA.
AC 004610;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Non-structural polypeptide [Contains: RNA-directed RNA polymerase
(EC 2.7.7.48); Helicase].
OS Hepatitis E virus (strain Myanmar) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=31769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93227573; PubMed=8470371;
RA Aye T.T., Uchida T., Ma M.Z., Iida F., Shikata T., Ichikawa M.,
RA Rikihisa T., Wino K.;
RT "Sequence and gene structure of the hepatitis E virus isolated from
RT Myanmar.";
RT Virus genes 7:95-109(1993).
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D10330; BA01172.1; -
CC PIR: A41.001; -
CC DR MEROPS: C41.001; -
CC DR InterPro: IPR002589; A1pp.
CC DR InterPro: IPR001230; Prenyl_site.
CC DR InterPro: IPR001788; RNA_dep_RNAPol2.
CC DR InterPro: IPR002588; V_methyltransf.
CC DR InterPro: IPR000606; Viral_helicase1.
CC DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
CC DR Pfam: PF01443; Viral_helicase1; 1.
CC DR Pfam: PF01660; Vmethyltransf; 1.
CC DR Pfam: PF01661; A1pp; 1.
CC DR SMART: SM00506; A1pp; 1.
CC KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
CC ATP-binding.
CC FT NP_BIND 975 982 ATP (POTENTIAL).
CC SEQUENCE 1693 AA; 185215 MW; AAB4C9140A7E21EA CRC64;

Query Match 27.2%; Score 46.5; DB 1; Length 1693;
Best Local Similarity 37.0%; Pred. No. 89;
Matches 10; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

QY 2 RTHRSMEIRTPDINPAMY-ASRGIRP 27
DB 904 RNHRGDELTYLPFLAARFEANRPTRP 930

RESULT 15
Y468_MYCGE STANDARD: PRT: 1783 AA.
AC 049460;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 35, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG468.
GN MG468.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,

[illegible]

Search completed: May 1, 2003, 14:33:19
Job time : 13.1719 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:13 ; Search time 15.0156 Seconds
(without alignments)
60.744 Million cell updates/sec

Title: US-09-446-543a-61

Perfect score: 171
Sequence: 1 SRTTHRSMEIRTPDINPAMYASRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

Issued Patents, AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCIRUS.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171	100.0	31	3	US-09-105-678A-9
2	171	100.0	31	3	US-09-105-678A-9
3	171	100.0	31	4	US-08-776-971-61
4	171	100.0	31	4	US-09-421-208-9
5	171	100.0	31	4	US-09-421-208-9
6	171	100.0	31	4	US-09-560-915-15
7	171	100.0	31	4	US-09-105-678A-44
8	171	100.0	32	3	US-08-776-971-62
9	171	100.0	32	4	US-09-421-208-44
10	171	100.0	33	3	US-09-105-678A-45
11	171	100.0	33	4	US-08-776-971-63
12	171	100.0	33	4	US-09-421-208-45
13	171	100.0	87	4	US-08-776-971-59
14	171	100.0	87	4	US-08-776-971-135
15	171	100.0	87	4	US-08-776-971-138
16	158	92.4	31	3	US-09-105-678A-7
17	158	92.4	31	3	US-09-105-678A-31
18	158	92.4	31	4	US-08-776-971-5
19	158	92.4	31	4	US-08-776-971-97
20	158	92.4	31	4	US-09-421-208-7
21	158	92.4	31	4	US-09-421-208-31
22	158	92.4	31	4	US-09-560-915-13
23	158	92.4	32	3	US-09-105-678A-32
24	158	92.4	32	4	US-08-776-971-6
25	158	92.4	32	4	US-09-421-208-32
26	158	92.4	33	3	US-09-105-678A-33
27	158	92.4	33	4	US-08-776-971-7

28	158	92.4	33	4	US-09-421-208-33	Sequence 33, Appl
29	158	92.4	98	4	US-08-776-971-1	Sequence 1, Appl
30	158	92.4	98	4	US-08-776-971-44	Sequence 44, Appl
31	158	92.4	98	4	US-08-776-971-122	Sequence 122, Appl
32	158	92.4	98	4	US-08-776-971-131	Sequence 131, Appl
33	158	92.4	98	4	US-08-776-971-136	Sequence 136, Appl
34	154	90.1	98	4	US-08-776-971-115	Sequence 115, App
35	154	90.1	31	3	US-09-105-678A-8	Sequence 8, Appl
36	149	87.1	31	3	US-09-105-678A-37	Sequence 37, Appl
37	149	87.1	31	4	US-09-172-353-4	Sequence 4, Appl
38	149	87.1	31	4	US-08-776-971-47	Sequence 47, Appl
39	149	87.1	31	4	US-09-421-208-8	Sequence 8, Appl
40	149	87.1	31	4	US-09-421-208-37	Sequence 37, Appl
41	149	87.1	31	4	US-09-560-915-14	Sequence 14, Appl
42	149	87.1	32	3	US-09-105-678A-38	Sequence 38, Appl
43	149	87.1	32	4	US-08-776-971-48	Sequence 48, Appl
44	149	87.1	32	4	US-09-421-208-38	Sequence 38, Appl
45	149	87.1	32	4	US-09-421-208-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-09-105-678A-9
Sequence 9, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-9

Query Match 100.0%; Score 171; DB 3; Length 31;
Best Local Similarity 100.0%; Pred. No. 7.1e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 SRTTHRSMEIRTPDINPAMYASRGIRPVGRF 31

Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

US-09-105-678A-43
; Sequence 43, Application US/09105678A
; Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS: linear

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-43

Query Match 100.0%; Score 171; DB 3; Length 31;

Best Local Similarity 100.0%; Pred. No. 7.1e-19;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

RESULT 3

US-08-776-971-61

; Sequence 61, Application US/08776971B

; Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

APPLICANT: Hinuma, Shuji

RESULT 4

US-09-421-208-9

; Sequence 9, Application US/09421208

; Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

APPLICANT: Suenaga, Masato

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 61:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 61:

US-08-776-971-61

Query Match 100.0%; Score 171; DB 4; Length 31;

Best Local Similarity 100.0%; Pred. No. 7.1e-19;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/421,208
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-9

Query Match
Best Local Similarity 100.0%; Score 171; DB 4; Length 31;
Pred. No. 7,1e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
|||||

RESULT 5
US-09-421-208-43
Sequence 43, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-43

Query Match
Best Local Similarity 100.0%; Score 171; DB 4; Length 31;
Pred. No. 7,1e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
|||||

RESULT 6
US-09-560-915-15
Sequence 15, Application US/09560915
Patent No. 6383764
GENERAL INFORMATION:
APPLICANT: Civeill, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Therapeutic Compositions and Methods
TITLE OF INVENTION: Relating To Prolactin Releasing Peptide (PrRP)
FILE REFERENCE: P-UC 3534
CURRENT APPLICATION NUMBER: US/09/560,915
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15
LENGTH: 31
TYPE: PRT
ORGANISM: Homo Sapien
US-09-560-915-15

Query Match
Best Local Similarity 100.0%; Score 171; DB 4; Length 31;
Pred. No. 7,1e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
|||||

RESULT 7
US-09-105-678A-44
Sequence 44, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A

FILED DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678a-44

Query Match 100.0%; Score 171; DB 3; Length 32;
Best Local Similarity 100.0%; Pred. No. 7.3e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETPTDINPAMYASRGIRPVGRF 31
DB 1 SRTHRSMETPTDINPAMYASRGIRPVGRF 31

RESULT 8

US-08-776-971-62
Sequence 62, Application US/08776971B
Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-08-776-971-62

Query Match 100.0%; Score 171; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 7.3e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETPTDINPAMYASRGIRPVGRF 31
DB 1 SRTHRSMETPTDINPAMYASRGIRPVGRF 31

RESULT 9

US-09-421-208-44
Sequence 44, Application US/09421208
Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-44

Query Match 100.0%; Score 171; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 7.3e-19;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

RESULT 10

US-09-105-678A-45
; Sequence 45, Application US/09105678A
; Patent No. 6103882
; GENERAL INFORMATION:
; APPLICANT: Suenaga, Masato
; APPLICANT: Moriya, Takeo
; APPLICANT: Tanaka, Yoko
; APPLICANT: Nishimura, Osamu
; TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,678A
; FILING DATE: 26-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 172118/1997
; FILING DATE: 27-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Conlin, David G.
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 48466-342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-105-678A-45

Query Match 100.0%; Score 171; DB 3; Length 33;
Best Local Similarity 100.0%; Pred. No. 7.6e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

RESULT 11

US-08-776-971-63
; Sequence 63, Application US/08776971B
; Patent No. 6228984
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Hadata, Yugo
; APPLICANT: Kawamata, Yuji
; APPLICANT: Hosoya, Masaki
; APPLICANT: Fujii, Ryo
; APPLICANT: Fukusumi, Shoji
; APPLICANT: Kitada, Chieko

TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 63:
US-08-776-971-63

Query Match 100.0%; Score 171; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 7.6e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31
Db 1 SRTHRSMETRPDINPAMYASRGIRPVGRF 31

RESULT 12

US-09-421-208-45
; Sequence 45, Application US/09421208
; Patent No. 6258561
; GENERAL INFORMATION:
; APPLICANT: Suenaga, Masato
; APPLICANT: Moriya, Takeo
; APPLICANT: Tanaka, Yoko
; APPLICANT: Nishimura, Osamu
; TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-45

Query Match 100.0%; Score 171; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 7,6e-19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRPDINPAMVYASRGIRPVGRF 31
Db 1 SRTHRSMETRPDINPAMVYASRGIRPVGRF 31

RESULT 13
US-08-776-971-59
Sequence 59, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Hinata, Yugo
Kawamata, Yui
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 87 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 59:
US-08-776-971-59

Query Match 100.0%; Score 171; DB 4; Length 87;
Best Local Similarity 100.0%; Pred. No. 2,4e-18;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SRTHRSMETRPDINPAMVYASRGIRPVGRF 31
Db 23 SRTHRSMETRPDINPAMVYASRGIRPVGRF 53

RESULT 14
US-08-776-971-135
Sequence 135, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Hinata, Yugo
Kawamata, Yui
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573
 FILING DATE: 18-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Conlin, David G.
 REGISTRATION NUMBER: 27,026
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-523-3400
 TELEFAX: 617-523-6440
 INFORMATION FOR SEQ ID NO: 135:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 87 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 SEQUENCE DESCRIPTION: SEQ ID NO: 135:
 US-08-776-971-135

Query Match 100.0%; Score 171; DB 4; Length 87;
 Best Local Similarity 100.0%; Pred. No. 2,4e-18;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTHRSMETRTPDINPAMYASRGIRPVGRF 31
 |||
 Db 23 SRTHRSMETRTPDINPAMYASRGIRPVGRF 53

RESULT 15
 US-08-776-971-138
 ; Sequence 138, Application US/0876971B
 ; Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji
 Habata, Yugo
 Kawamata, Yuji
 Hosoya, Masaki
 Fujii, Ryo
 Fukusumi, Shoji
 Kitada, Chieko
 TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
 NUMBER OF SEQUENCES: 140
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
 STREET: 130 Water Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/776, 971B
 FILING DATE: 06-Feb-1997
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP96/03821
 FILING DATE: 28-DEC-1996
 APPLICATION NUMBER: JP 7/343371
 FILING DATE: 28-DEC-1995
 APPLICATION NUMBER: JP 8/59419
 FILING DATE: 15-MAR-1996
 APPLICATION NUMBER: JP 8/211805
 FILING DATE: 12-AUG-1996
 APPLICATION NUMBER: JP 8/246573
 FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:
 NAME: Conlin, David G.
 REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-523-3400
 TELEFAX: 617-523-6440
 INFORMATION FOR SEQ ID NO: 138:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 87 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 SEQUENCE DESCRIPTION: SEQ ID NO: 138:
 US-08-776-971-138

Query Match 100.0%; Score 171; DB 4; Length 87;
 Best Local Similarity 100.0%; Pred. No. 2,4e-18;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTHRSMETRTPDINPAMYASRGIRPVGRF 31
 |||
 Db 23 SRTHRSMETRTPDINPAMYASRGIRPVGRF 53

Search completed: May 1, 2003, 14:33:54
 Job time: 16.0156 secs

1
2
3
4
5

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:46:12 ; Search time 15 Seconds
(without alignments)
64,730 Million cell updates/sec

Title: SEQ74-PLUS-73
Perfect score: 156
Sequence: 1 SRXHXSMEXKTPDINPAMWYXXRGIRPVGRFX 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by the chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	94.2	31	3	US-09-105-678A-8
2	147	94.2	31	3	US-09-105-678A-9
3	147	94.2	31	3	US-09-105-678A-37
4	147	94.2	31	3	US-09-105-678A-43
5	147	94.2	31	4	US-09-172-353-4
6	147	94.2	31	4	US-08-776-971-47
7	147	94.2	31	4	US-08-776-971-61
8	147	94.2	31	4	US-09-421-208-8
9	147	94.2	31	4	US-09-421-208-9
10	147	94.2	31	4	US-09-421-208-37
11	147	94.2	31	4	US-09-421-208-43
12	147	94.2	31	4	US-09-560-915-14
13	147	94.2	31	4	US-09-560-915-15
14	147	94.2	32	3	US-09-105-678A-38
15	147	94.2	32	3	US-09-105-678A-44
16	147	94.2	32	4	US-08-776-971-48
17	147	94.2	32	4	US-08-776-971-62
18	147	94.2	32	4	US-09-421-208-38
19	147	94.2	32	4	US-09-421-208-44
20	147	94.2	33	3	US-09-105-678A-39
21	147	94.2	33	3	US-09-105-678A-45
22	147	94.2	33	4	US-08-776-971-49
23	147	94.2	33	4	US-08-776-971-63
24	147	94.2	33	4	US-09-421-208-39
25	147	94.2	33	4	US-09-421-208-45
26	147	94.2	83	4	US-08-776-971-45
27	147	94.2	83	4	US-08-776-971-124

28	147	94.2	83	4	US-08-776-971-137	Sequence 137, App
29	147	94.2	87	4	US-08-776-971-59	Sequence 59, Appl
30	147	94.2	87	4	US-08-776-971-135	Sequence 135, App
31	147	94.2	87	4	US-08-776-971-138	Sequence 138, App
32	146	93.6	31	3	US-09-105-678A-7	Sequence 7, Appl
33	146	93.6	31	3	US-09-105-678A-31	Sequence 31, Appl
34	146	93.6	31	4	US-08-776-971-5	Sequence 5, Appl
35	146	93.6	31	4	US-08-776-971-97	Sequence 97, Appl
36	146	93.6	31	4	US-09-421-208-31	Sequence 31, Appl
37	146	93.6	31	4	US-09-421-208-7	Sequence 7, Appl
38	146	93.6	31	4	US-09-560-915-13	Sequence 13, Appl
39	146	93.6	32	3	US-09-105-678A-32	Sequence 32, Appl
40	146	93.6	32	4	US-08-776-971-6	Sequence 6, Appl
41	146	93.6	32	4	US-09-421-208-32	Sequence 32, Appl
42	146	93.6	33	3	US-09-105-678A-33	Sequence 33, Appl
43	146	93.6	33	4	US-08-776-971-7	Sequence 7, Appl
44	146	93.6	33	4	US-09-421-208-33	Sequence 33, Appl
45	146	93.6	98	4	US-08-776-971-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-105-678A-8
Sequence 8, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conliff, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-8
Query Match 94.2%; Score 147; DB 3; Length 31;
Best Local Similarity 83.9%; Pred. No. 1,2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 SRXHXSMEXKTPDINPAMWYXXRGIRPVGRFX 31

DB 1 SRAHSHMETRTPDINPAWYTGRCIRPVGRF 31

RESULT 2

US-09-105-678A-9

Sequence 9, Application US/09105678A

Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-9

Query Match 94.2%; Score 147; DB 3; Length 31;

Best Local Similarity 83.9%; Pred. No. 1.2e-12;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRAHSHMETRTPDINPAWYTGRCIRPVGRF 31

DB 1 SRAHSHMETRTPDINPAWYTGRCIRPVGRF 31

RESULT 3

US-09-105-678A-37

Sequence 37, Application US/09105678A

Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-37

Query Match 94.2%; Score 147; DB 3; Length 31;

Best Local Similarity 83.9%; Pred. No. 1.2e-12;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRAHSHMETRTPDINPAWYTGRCIRPVGRF 31

DB 1 SRAHSHMETRTPDINPAWYTGRCIRPVGRF 31

RESULT 4

US-09-105-678A-43

Sequence 43, Application US/09105678A

Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo

APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-43

Query Match 94.2%; Score 147; DB 3; Length 31;

Best Local Similarity 83.9%; Pred. No. 1.2e-12;

Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRAHSHMETRTPDINPAWYTGRCIRPVGRF 31

DB 1 SRAHSHMETRTPDINPAWYTGRCIRPVGRF 31

TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-43

Query Match
Best Local Similarity 94.2%; Score 147; DB 3; Length 31;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRHRHSMETRPDINPAMYASRGIRPVGRF 31

RESULT 5
US-09-172-353-4
Sequence 4, Application US/09172353
Patent No. 6197530
GENERAL INFORMATION:
APPLICANT: Stricker-Kongra, Alain
APPLICANT: Gu, Wei
TITLE OF INVENTION: GPR10 AS A TARGET FOR IDENTIFYING WEIGHT MODULATING COMPOUNDS
FILE REFERENCE: 07334/102001
CURRENT APPLICATION NUMBER: US/09/172,353
CURRENT FILING DATE: 1998-10-14
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 31
TYPE: PRT
ORGANISM: Mus musculus
US-09-172-353-4

Query Match
Best Local Similarity 94.2%; Score 147; DB 4; Length 31;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRAHSHMETRPDINPAMYTGIRPVGRF 31

RESULT 6
US-08-776-971-47
Sequence 47, Application US/0876971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fujii, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-08-776-971-47

Query Match
Best Local Similarity 94.2%; Score 147; DB 4; Length 31;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRAHSHMETRPDINPAMYTGIRPVGRF 31

RESULT 7
US-08-776-971-61
Sequence 61, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fujii, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 61:
US-08-776-971-61

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHXSMEXRTPDINPANYXXRGIRPVGRF 31
Db 1 STHHSHMETRTPDINPANYASRGIRPVGRF 31

RESULT 8
US-09-421-208-8
Sequence 8, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-8

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHXSMEXRTPDINPANYXXRGIRPVGRF 31
Db 1 SRAHSHMETRTPDINPANYTGRGIRPVGRF 31

RESULT 9
US-09-421-208-9
Sequence 9, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-9

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SRXHXSMEXRTPDINPANYXXRGIRPVGRF 31

DB 1 SRTXHSMEIRTPDINPAWYASRGIRPVGRF 31

RESULT 10

US-09-421-208-37
Sequence 37, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-37

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHSMEIRTPDINPAWYASRGIRPVGRF 31
DB 1 SRAHSHMETRTPDINPAWYTGIRPVGRF 31

RESULT 11

US-09-421-208-43
Sequence 43, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-43

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHSMEIRTPDINPAWYASRGIRPVGRF 31
DB 1 SRTXHSMEIRTPDINPAWYASRGIRPVGRF 31

RESULT 12

US-09-560-915-14
Sequence 14, Application US/09560915
Patent No. 6383764
GENERAL INFORMATION:
APPLICANT: Clivell, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Therapeutic Compositions and Methods
TITLE OF INVENTION: Relating To Proclatin Releasing Peptide (PRRP)
FILE REFERENCE: P-UC 3534
CURRENT APPLICATION NUMBER: US/09/560,915
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 31
TYPE: PRT
ORGANISM: Rattus
US-09-560-915-14

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHSMEIRTPDINPAWYASRGIRPVGRF 31
DB 1 SRAHSHMETRTPDINPAWYTGIRPVGRF 31

RESULT 13

US-09-560-915-15
; Sequence 15, Application US/09560915
; Patent No. 6383764
; GENERAL INFORMATION:
; APPLICANT: Civalli, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Therapeutic Compositions and Methods
; FILE REFERENCE: P-UC 3534
; CURRENT APPLICATION NUMBER: US/09/560,915
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-560-915-15

Query Match 94.2%; Score 147; DB 4; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRHXHSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRTHRHSMETRPDINPAMYASRGIRPVGRF 31

RESULT 14
US-09-105-678A-38
; Sequence 38, Application US/09105678A
; Patent No. 6103882
; GENERAL INFORMATION:
; APPLICANT: Suenaga, Masato
; APPLICANT: Moriya, Takeo
; APPLICANT: Tanaka, Yoko
; APPLICANT: Nishimura, Osamu
; TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,678A
; FILING DATE: 26-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 172118/1997
; FILING DATE: 27-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Conlin, David G.
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 48466-342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-105-678A-38

Query Match 94.2%; Score 147; DB 3; Length 32;
Best Local Similarity 83.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRHXHSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRTHRHSMETRPDINPAMYASRGIRPVGRF 31

RESULT 15
US-09-105-678A-44
; Sequence 44, Application US/09105678A
; Patent No. 6103882
; GENERAL INFORMATION:
; APPLICANT: Suenaga, Masato
; APPLICANT: Moriya, Takeo
; APPLICANT: Tanaka, Yoko
; APPLICANT: Nishimura, Osamu
; TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,678A
; FILING DATE: 26-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 172118/1997
; FILING DATE: 27-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Conlin, David G.
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 48466-342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-105-678A-44

Query Match 94.2%; Score 147; DB 3; Length 32;
Best Local Similarity 83.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRHXHSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRTHRHSMETRPDINPAMYASRGIRPVGRF 31

Search completed: May 1, 2003, 14:49:48
Job time : 15 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 12.0312 Seconds

(without alignments)
188.386 Million cell updates/sec

Title: US-09-446-543a-74

Perfect score: 47

Sequence: 1 SRXHXSMEXR 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

SPTREMBL.21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioplasmid:*
17: sp_archaeoplasmid:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	89.4	98	6	Q8WN12
2	35	74.5	450	4	Q95367
3	35	74.5	451	4	Q12804
4	34	72.3	672	16	Q8XBC4
5	33	70.2	340	2	Q9RHI9
6	33	70.2	4133	13	Q9DE12
7	33	70.2	4134	13	Q8QGX4
8	32	68.1	196	17	Q9HNP4
9	32	68.1	277	11	Q9DRA2
10	32	68.1	286	11	Q9DAR5
11	32	68.1	383	16	Q84156
12	32	68.1	450	11	Q9DA19
13	32	68.1	648	16	Q9XOW5
14	32	68.1	712	6	Q8Y2S0
15	31	66.0	141	6	Q8WMO1
16	31	66.0	165	2	Q9JN22

17	31	66.0	685	5	Q8T016	Q8T016 drosophila
18	31	66.0	756	5	Q9U116	Q9U116 drosophila
19	31	66.0	800	5	Q9V731	Q9V731 drosophila
20	31	66.0	905	5	Q9W178	Q9W178 drosophila
21	30	63.8	147	5	Q9B168	Q9B168 drosophila
22	30	63.8	274	10	Q949G1	Q949G1 oryza sativ
23	30	63.8	330	16	Q9K7F3	Q9K7F3 bacillus ha
24	30	63.8	330	16	Q9A0B5	Q9A0B5 streptococc
25	30	63.8	332	16	Q05268	Q05268 bacillus su
26	30	63.8	387	16	Q8XXK0	Q8XXK0 ralsstonia s
27	30	63.8	393	5	Q23359	Q23359 caenorhabdi
28	30	63.8	394	10	Q9ASB8	Q9ASB8 oryza sativ
29	30	63.8	462	11	Q9CY07	Q9CY07 mus musculu
30	30	63.8	509	16	Q9XA23	Q9XA23 streptomyce
31	30	63.8	528	3	Q9C2S0	Q9C2S0 neurospora
32	30	63.8	599	16	Q9K4U0	Q9K4U0 vibrio chol
33	30	63.8	707	10	Q9AUS0	Q9AUS0 oryza sativ
34	30	63.8	1889	10	Q94H07	Q94H07 oryza sativ
35	29	61.7	96	8	Q94Y20	Q94Y20 pyllaetella l
36	29	61.7	122	16	Q9K0W8	Q9K0W8 neisseria m
37	29	61.7	122	16	Q9J710	Q9J710 neisseria m
38	29	61.7	147	17	Q8RPV9	Q8RPV9 methanosarc
39	29	61.7	152	10	Q8S5H7	Q8S5H7 oryza sativ
40	29	61.7	160	11	Q91VZ5	Q91VZ5 mus musculu
41	29	61.7	177	11	Q8Q2T3	Q8Q2T3 mus musculu
42	29	61.7	194	10	Q41320	Q41320 secale cere
43	29	61.7	220	12	Q9Q3U4	Q9Q3U4 cymbidium m
44	29	61.7	233	3	Q9P826	Q9P826 candida alb
45	29	61.7	336	10	Q9AV29	Q9AV29 oryza sativ

ALIGNMENTS

RESULT 1

ID Q8WN12 PRELIMINARY; PRT; 98 AA.
AC Q8WN12;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DI 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Proprolactin-releasing peptide.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Curlewis J.D., Kusters D.H.L., Barclay J.L., Anderson S.T.;
RT "Prolactin-releasing peptide (PrRP) in the ewe: cDNA cloning, mRNA
RT distribution and effects on prolactin secretion in vitro and in
RT vivo".
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF450453; ALU47178.1;
SO SEQUENCE 98 AA; 10513 MW; 2A53331ED62CAAB5 CRC64;

Query Match 89.4%; Score 42; DB 6; Length 98;
Best Local Similarity 72.7%; Pred. No. 0.072;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 23 SRXHXSMEXR 33
RESULT 2
ID Q95367 PRELIMINARY; PRT; 450 AA.
AC Q95367;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

```

DE CBF1 interacting corepressor CIR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99093483; PubMed=9874765;
RA Hsieh J.J., Zhou S., Chen L., Young D.B., Hayward S.D.;
RT "CIR, a corepressor linking the DNA binding factor CBF1 to the histone
RT deacetylase complex."
RL Proc. Natl. Acad. Sci. U.S.A. 96:23-28(1999).
DR EMBL, AF098297; AAD05243.1;
SQ SEQUENCE 450 AA; 52328 MW; 96AA8BF69CB73357 CRC64;

Query Match
Best Local Similarity 74.5%; Score 35; DB 4; Length 450;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 359 SRTXKHSPEKR 369

RESULT 3
ID 012804 PRELIMINARY; PRT; 451 AA.
AC 012804;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE Receptin.
GN RECEPTIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Chai K.X., Li L., Chao J., Chao L.;
RT "Receptin: a novel human liver cDNA encoding a serpin-like molecule."
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; U03644; AAA17853.1;
DR InterPro; IPR000345; Cytochrome_b.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.1.
SQ SEQUENCE 451 AA; 52942 MW; 157F04804DA0817E CRC64;

Query Match
Best Local Similarity 74.5%; Score 35; DB 4; Length 451;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 360 SRTXKHSPEKR 370

RESULT 4
ID 08XBC4 PRELIMINARY; PRT; 672 AA.
AC 08XBC4;
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Hydrogenase 4 membrane subunit.
GN HYTB OR 23742 OR ECS3344.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

```

```

RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIND 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005478; AAG57592.1;
DR EMBL; AF002561; BAB36767.1;
DR InterPro; IPR003918; NADHdb_oxred4.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; Oxidored_q1.
DR PRINTS; PR01437; NUOXDRDTASE4.
SQ SEQUENCE 672 AA; 72659 MW; D983361A569AC70E CRC64;

Query Match
Best Local Similarity 72.3%; Score 34; DB 16; Length 672;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 9
Db 364 SRTXKHSPEKR 372

RESULT 5
ID 09RH19 PRELIMINARY; PRT; 340 AA.
AC 09RH19;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Cytochrome oxidase D subunit B.
GN CYDB.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RA Lee J.S., Kang H.S.;
RT "Sequence analysis of 6563 cosmid clone of Zymomonas mobilis ZM4
RT containing rna operon."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF088897; AAF18273.1;
DR InterPro; IPR003317; Cyto_ox_2.
DR Pfam; PF02322; Cyto_ox_2.
DR TIGRPFAM; TIGR00203; cydB; 1.
SQ SEQUENCE 340 AA; 37206 MW; 3D7038A3F72EC316 CRC64;

Query Match
Best Local Similarity 70.2%; Score 33; DB 2; Length 340;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 SRXHXSMEXR 11
Db 107 RVXAHKSEAR 116

```

```

RESULT 6
Q9DEI2 PRELIMINARY; PRT; 4133 AA.
AC Q9DEI2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DNA-dependent protein kinase catalytic subunit.
GN DNA-PKCS.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LYMPHOCYTE;
RX MEDLINE=20456723; PubMed=11003390;
RA Fujimori A., Tatsumi K., Abe M.;
RA Kawahara A., Tatsumi K., Fukunura R., Ohnata T., Takahashi H.,
RT Identification of four highly conserved regions in DNA-PKcs.;
RL EMBL; AB016240; BAA36956.1; -
DR InterPro: IPR003151; FAT.
DR InterPro: IPR003152; FATC.
DR InterPro: IPR000403; PI3_PI4_kinase.
DR Pfam: PF02259; FAT; 1.
DR Pfam: PF02260; FATC; 1.
DR Pfam: PF00454; PI3_PI4_kinase; 1.
DR SMART: SM00146; PI3K; 1.
DR PROSITE: PS00915; PI3_4_KINASE_1; UNKNOWN_1.
DR PROSITE: PS0290; PI3_4_KINASE_3; 1.
KW kinase.
SQ SEQUENCE 4133 AA; 472522 MW; 0A15C6131466A068 CRC64;

Query Match
Best Local Similarity 70.2%; Score 33; DB 13; Length 4133;
Best Local Similarity 75.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RXHXSME 9
Db 1766 RDRHSME 1773

RESULT 7
Q8GKX4 PRELIMINARY; PRT; 4134 AA.
AC Q8GKX4;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE DNA-dependent protein kinase catalytic subunit.
GN PRKDC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21829525; PubMed=11839092;
RA Fujimori A., Hashimoto H., Araki R., Saito T., Sato S., Kasama Y.,
RA Tetsunari Y., Mori M., Fukunura R., Ohnata T., Tatsumi K., Abe M.;
RT Sequence Analysis of 193.4 and 83.9 kbp of Mouse and Chicken Genomic
RT DNAs Containing the Entire Prkdc (DNA-PKcs) Gene.;
RL Radiat. Res. 157:298-305(2002).
DR EMBL; AB028136; BAB91148.1; -
KW kinase.
SQ SEQUENCE 4134 AA; 472647 MW; 8907B1EA6E7E9A8 CRC64;

Query Match
Best Local Similarity 70.2%; Score 33; DB 13; Length 4134;
Best Local Similarity 75.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 RXHXSME 9
Db 1767 RDRHSME 1774

RESULT 8
Q9HNP4 PRELIMINARY; PRT; 196 AA.
AC Q9HNP4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Vng2006C.
GN Vng2006C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Ballga N.S., Thorsson V., Sirogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weltl R., Goo Y.A.,
RA Leitbauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angvine C.M., Dale H.,
RA Isenbarger T.A., Beck R.F., Pohlschoder M., Spudich J.L., Jung K.-H.,
RA Alan M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AF005094; AAG20176.1; -
KW Complete proteome.
SQ SEQUENCE 196 AA; 21450 MW; 2F97F424E7C9E73 CRC64;

Query Match
Best Local Similarity 68.1%; Score 32; DB 17; Length 196;
Best Local Similarity 54.5%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 SRXHSMEXR 11
Db 112 SRHQSLEDR 122

RESULT 9
Q9D8A2 PRELIMINARY; PRT; 277 AA.
AC Q9D8A2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 1700023B02R1k protein.
GN 1700023B02R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kato K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaesteland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita K., Gardiola M.,

```

RA Gustinlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.",
 RL Nature 409:685-690(2001).
 DR EMBL: AK008242; BAF25553.1; -
 DR MGD: MGI:1914185; 1700023B02RIK.
 DR PRINTS: PRO1574; TUBBYPROTEIN.
 SQ SEQUENCE 277 AA; 31797 MW; 4BE4D95516FBAC32 CRC64;

Query Match 68.1%; Score 32; DB 11; Length 277;
 Best Local Similarity 54.5%; Pred. No. 31;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
 DB 200 SRSHRSPERK 210

RESULT 10
 ID Q9DAE5 PRELIMINARY; PRT; 286 AA.
 AC Q9DAE5;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 1700023B02RIK protein.
 GN 1700023B02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.",
 RL Nature 409:685-690(2001).
 DR EMBL: AK005900; BAB24307.1; -
 DR MGD: MGI:1914185; 1700023B02RIK.
 DR PRINTS: PRO1574; TUBBYPROTEIN.
 SQ SEQUENCE 286 AA; 32818 MW; 068003C5E894827B CRC64;

Query Match 68.1%; Score 32; DB 11; Length 286;
 Best Local Similarity 54.5%; Pred. No. 32;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
 DB 209 SRSHRSPERK 219

RESULT 11
 ID 084156 PRELIMINARY; PRT; 383 AA.
 AC 084156;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Phospholipase D endonuclease superfamily.
 GN Cris14.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/CX;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.,
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis.",
 RL Science 282:754-759(1998).
 DR EMBL: AE001289; AAC67745.1; -
 DR InterPro: IPR001736; PLD.
 DR Pfam: PF00614; PLDC; 2.
 DR SMART: SM00155; PLDC; 2.
 KW Endonuclease; Complete proteome.
 SQ SEQUENCE 383 AA; 43383 MW; 4BF73A165353D2FA CRC64;

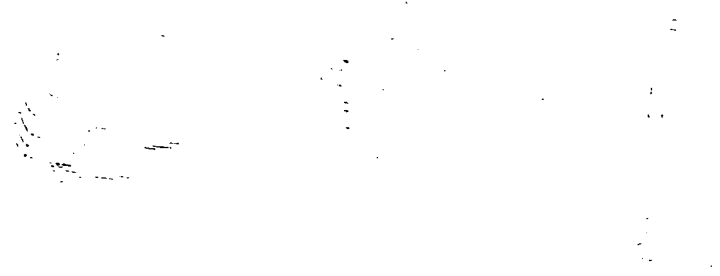
Query Match 68.1%; Score 32; DB 16; Length 383;
 Best Local Similarity 54.5%; Pred. No. 43;
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
 DB 18 SRSHRSPERK 28

RESULT 12
 ID Q9DA19 PRELIMINARY; PRT; 450 AA.
 AC Q9DA19;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 1700023B02RIK protein.
 GN 1700023B02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,

RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK006260; BAB2448.1;
 DR MGD: MGI:1914185; 170023B02R1K.
 DR InterPro: IPR000345; CytC_heme_bind.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 SQ SEQUENCE 450 AA; 51852 MW; F32P11BE6DA4EAC CRC64;
 Query Match 68.1%; Score 32; DB 11; Length 450;
 Best Local Similarity 54.5%; Pred. No. 50;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 SRXHXSMEXR 11
 DB 373 SRSHSHSPERK 383
 RESULT 13
 ID Q9XOW5 PRELIMINARY; PRT; 648 AA.
 AC Q9XOW5;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 17, Last sequence update)
 DE ATP-dependent DNA helicase.
 GN TM1238.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwyn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Lither K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima."
 RL Nature 399:323-329(1999).
 DR EMBL: AF001779; AAD36313.1;
 DR HSSP: P56255; 1PJR.
 DR TIGR: TM1238;
 DR InterPro: IPR000212; UvrD-helicase.
 DR Pfam: PF00580; UvrD-helicase; 1.
 KW Helicase; Complete proteome.
 SQ SEQUENCE 648 AA; 75842 MW; 4274FFB319250280 CRC64;
 Query Match 68.1%; Score 32; DB 16; Length 648;
 Best Local Similarity 62.5%; Pred. No. 72;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 RXHXHSM 9
 DB 370 RSHSHSL 377
 RESULT 14
 ID Q8Y2S0 PRELIMINARY; PRT; 712 AA.
 AC Q8Y2S0;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DE Probable acyl-CoA carboxylase alpha chain protein.
 GN RSC0265 OR RS05771.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Beta subdivision; Ralstonia group;
 OX NCBI_TaxID=305;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoudat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlet M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chaudler C., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigler P., Thebaud P., Whalen M., Winkler P., Levy M.,
 RA Weissbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
 RL Nature 415:497-502(2002).
 DR EMBL: AL646058; CAD13793.1;
 DR InterPro: IPR000089; Biotin_lipoyl.
 DR InterPro: IPR000901; CPSase.
 DR Pfam: PF02785; Biotin_carb_C; 1.
 DR Pfam: PF00364; Biotin_lipoyl; 1.
 DR Pfam: PF00289; CPSase_L_D2; 1.
 DR Pfam: PF02786; CPSase_L_D2; 1.
 DR PROSITE: PS00866; CPSASE_1; 1.
 DR PROSITE: PS00867; CPSASE_2; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 712 AA; 76240 MW; 37D79D7B8C222797 CRC64;
 Query Match 68.1%; Score 32; DB 16; Length 712;
 Best Local Similarity 50.0%; Pred. No. 80;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 RXHXHSMEXR 11
 DB 366 RIHGHAEAR 375
 RESULT 15
 ID Q8WM01 PRELIMINARY; PRT; 141 AA.
 AC Q8WM01;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DE Transformer-2 alpha (Fragment).
 GN Macaca mulatta (Rhesus macaque).
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISUB-HYPOTHALAMUS;
 RA Mungenast A.E., Ojeda S.R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY072878; AAL68856.1;
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 141 AA; 15974 MW; A4B9A50C50670CA5 CRC64;
 Query Match 66.0%; Score 31; DB 6; Length 141;
 Best Local Similarity 54.5%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 SRXHXSMEXR 11
 DB 50 SRSHSHSHRR 60
 Search completed: May 1, 2003, 14:33:01
 Job time : 16.0312 secs



GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 3.6038 Seconds

(without alignments)
126.404 Million cell updates/sec

Title: US-09-446-543a-74

Perfect score: 47

Sequence: 1 SRXHXSMEXR 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	91.5	83	1	PRRP_RAT
2	42	89.4	87	1	PRRP_HUMAN
3	42	89.4	98	1	PRRP_BOVIN
4	34	72.3	672	1	HYER_ECOLI
5	33	70.2	497	1	YH80_MYCOU
6	32	68.1	452	1	YH4_YEAST
7	32	68.1	460	1	VP41_BPAPS
8	31	66.0	277	1	MCRA_ECOLI
9	31	66.0	282	1	TR2A_HUMAN
10	31	66.0	495	1	HSP1_ARATH
11	30	63.8	333	1	ILVC_MYCAV
12	30	63.8	333	1	ILVC_MYCTU
13	30	63.8	340	1	ILVC_CAMBE
14	30	63.8	772	1	SM3A_RAT
15	30	63.8	1224	1	MSN5_YEAST
16	29	61.7	145	1	OAZ_ONCVO
17	29	61.7	153	1	ITX1_MOUSE
18	29	61.7	219	1	LOL2_MOUSE
19	29	61.7	405	1	IPSP_MOUSE
20	29	61.7	462	1	US29_HCMVA
21	29	61.7	513	1	REP_HUMAN
22	29	61.7	522	1	REP_MOUSE
23	29	61.7	774	1	LOL2_HUMAN
24	29	61.7	915	1	Y007_CAEBL
25	29	61.7	2869	1	RBP1_PLAIV
26	28	59.6	277	1	T2FB_DROME
27	28	59.6	319	1	TRXB_BUCAP
28	28	59.6	331	1	UL38_HCMVA
29	28	59.6	338	1	ILVC_CORGL
30	28	59.6	340	1	ILVC_STPRR
31	28	59.6	363	1	ODPA_SMIMA
32	28	59.6	388	1	ODPT_HUMAN
33	28	59.6	389	1	ODPA_ARATH

34	28	59.6	389	1	ODPA_PIG	P29804 sus scrofa
35	28	59.6	390	1	ODPA_HUMAN	P08559 homo sapien
36	28	59.6	390	1	ODPA_MOUSE	P35486 mus musculus
37	28	59.6	390	1	ODPA_RAT	P26284 rattus norv
38	28	59.6	391	1	ODPA_SOLITU	P32903 solanum tub
39	28	59.6	391	1	ODPT_MOUSE	P52487 mus musculus
40	28	59.6	391	1	ODPT_RAT	P06437 rattus norv
41	28	59.6	397	1	ODPA_CAEBL	P52893 caenorhabdi
42	28	59.6	397	1	ODPA_PEA	P52902 pisum sativ
43	28	59.6	507	1	C7C3_ARATH	064638 arabidopsis
44	28	59.6	521	1	IMDH_CHYVI	050316 chlorobium
45	28	59.6	576	1	SYR_YERPE	082ev7 yersinia pe

ALIGNMENTS

RESULT 1	PRRP_RAT	STANDARD:	PRT:	83 AA.
AC	P81278:			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Proactin-releasing peptide precursor (PRRP) (Proactin-releasing hormone) [Contains: Proactin-releasing peptide PRRP31; Proactin-releasing peptide PRRP20].			
GN	PRH.			
OS	Rattus norvegicus (Rat).			
OC	Euarystola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
NCBI_taxid:10116;				
NP	SEQUENCE FROM N.A.			
RP	TISSUE-BRAIN:			
RC	TISSUE-BRAIN:			
RA	MEDLINE-98268781; Pubmed-9607765;			
RA	Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,			
RA	Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,			
RA	Kurokawa T., Nishimura O., Onda H., Fujino M.;			
RA	"A proactin-releasing peptide in the brain.";			
RL	Nature 393:272-276(1998).			
RT	[2]			
RP	TISSUE SPECIFICITY.			
RX	MEDLINE-9942652; Pubmed-10498338;			
RA	Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S.,			
RA	Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H.,			
RA	Sunino Y., Fujino M.;			
RT	"Tissue distribution of proactin-releasing peptide (PRRP) and its receptor.";			
RL	Regul. Pept. 83:1-10(1999).			
CC	-I- FUNCTION: Stimulates proactin (PRP) release and regulates the expression of proactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRP.			
CC	-I- TISSUE SPECIFICITY: Widely expressed, with highest levels in medulla oblongata and hypothalamus.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; AB015418; BAA29026.1;			
KW	Hormone; Amidation; Signal; Cleavage on pair of basic residues.			
FT	SIGNAL.			
FT	PEPTIDE			
FT	PEPTIDE			
FT	MOD_RES			
SO	SEQUENCE			
Query Match	83 AA; 9215 MW; DOC75A264EEF4F29 CRC64;			
	91.5%; Score 43; DB 1; Length 83;			

Best Local Similarity 72.7%; Pred. No. 0.0078;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
DB 22 SRAHSHMEIR 32

RESULT 2

PRRP_HUMAN STANDARD; PRT; 87 AA.
ID PRRP_HUMAN
AC P81277;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Prolactin-releasing peptide precursor (PRP) (Prolactin-releasing hormone) [Contains: Prolactin-releasing peptide PRRP31; Prolactin-releasing peptide PRRP20].
DE hormone [Contains: Prolactin-releasing peptide PRRP31; Prolactin-releasing peptide PRRP20].
GN PRH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kikada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RL Nature 393:272-276(1998).
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=9942652; PubMed=10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S., Sekiguchi M., Kikada C., Kurokawa T., Nishimura O., Onda H., Sumino Y., Fujino M.;
RT "Tissue distribution of prolactin-releasing peptide (PRP) and its receptor.";
RL Regul. Pept. 83:1-10(1999).

CC -1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.
CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AB015419; BAA29027.1; -
DR MIM: 602663; -
KW Hormone; Amidation; Signal.
FT SIGNAL 1
FT PEPTIDE 23 BY SIMILARITY.
FT PEPTIDE 34 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT MOD_RES 53 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 53 AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ SEQUENCE 87 AA; 9639 MW; 22A2AF3F50CF981B CRC64;

Query Match 89.4%; Score 42; DB 1; Length 87;
Best Local Similarity 72.7%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
DB 23 SRAHSHMEIR 33

RESULT 3

PRRP_BOVIN STANDARD; PRT; 98 AA.
ID PRRP_BOVIN
AC P81264;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Prolactin-releasing peptide precursor (PRP) (Prolactin-releasing hormone) [Contains: Prolactin-releasing peptide PRRP31; Prolactin-releasing peptide PRRP20].
DE hormone [Contains: Prolactin-releasing peptide PRRP31; Prolactin-releasing peptide PRRP20].
GN PRH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX NCBI_TaxID=9913;

RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
RC TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S., Kikada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M., Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RL Nature 393:272-276(1998).
CC -1- FUNCTION: Stimulates prolactin (PRL) release and regulates the expression of prolactin through its receptor GPR10. May stimulate lactotrophs directly to secrete PRL.
CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AB015417; BAA29025.1; -
KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
FT SIGNAL 1
FT PEPTIDE 23
FT PEPTIDE 33 PROLACTIN-RELEASING PEPTIDE PRRP31.
FT MOD_RES 53 PROLACTIN-RELEASING PEPTIDE PRRP20.
FT MOD_RES 53 AMIDATION (G-54 PROVIDE AMIDE GROUP).
SQ SEQUENCE 98 AA; 10544 MW; 08AC35M13B0FA908 CRC64;

Query Match 89.4%; Score 42; DB 1; Length 98;
Best Local Similarity 72.7%; Pred. No. 0.015;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
DB 23 SRAHSHMEIR 33

RESULT 4

HYFB_ECOLI STANDARD; PRT; 672 AA.
ID HYFB_ECOLI
AC P23482;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrogenase-4 component B (EC 1.1.1.1).
GN HYFB OR B2482.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;

RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN-K12;
RA Andrews S.C., McClay J., Ambler A., Quail M., Berks B.C., Guest J.R.;
RL Submitted (Oct-1996) to the EMBL/Genbank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blatner F.R., Plunkett G., Iii, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97349980; PubMed-9205837;
 RA Yamanoto Y., Alida H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsunashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubraman S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horinouchi T.,
 RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 RN [4]
 RP SEQUENCE OF 1-34 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-91202105; PubMed-2016588;
 RA Andrews S.C., Harrison P.M., Guest J.R.,
 RT "A molecular analysis of the 53.3 minute region of the *Escherichia*
 RT coli linkage map".
 RL J. Gen. Microbiol. 137:361-367(1991).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (By similarity).
 CC SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
 CC OF CHLOROPLASTS OR MITOCHONDRIA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M63654; AAB86564.1;
 DR EMBL: AE000335; AAC75535.1;
 DR EMBL: D90876; BAA16360.1;
 DR EMBL: D90877; BAA16370.1;
 DR PIR: D49749; DA9749.
 DR Ecogene: EG11282; hyfB.
 DR InterPro: IPR001750; Oxidored_q1.
 DR Pfam: PF00361; Oxidored_q1; 1.
 KW Oxidoreductase; Transmembrane; Inner membrane; Complete proteome.
 FT TRANSMEM 31 26 POTENTIAL.
 FT TRANSMEM 51 51 POTENTIAL.
 FT TRANSMEM 80 100 POTENTIAL.
 FT TRANSMEM 120 140 POTENTIAL.
 FT TRANSMEM 165 185 POTENTIAL.
 FT TRANSMEM 200 220 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 312 332 POTENTIAL.
 FT TRANSMEM 343 363 POTENTIAL.
 FT TRANSMEM 385 405 POTENTIAL.
 FT TRANSMEM 428 448 POTENTIAL.
 FT TRANSMEM 475 495 POTENTIAL.
 FT TRANSMEM 505 525 POTENTIAL.
 FT TRANSMEM 532 552 POTENTIAL.
 FT TRANSMEM 652 672 POTENTIAL.
 SQ SEQUENCE 672 AA; 72582 MW; 3CAB82BD25E54B1B CRC64;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SRXHXSMEX 9
 Db 364 SRLHSHDME 372
 RESULT 5
 Y80 MYCTU
 ID Y80 MYCTU STANDARD: PRT: 497 AA.
 AC 006343;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV3480C
 GN RV3480C OR M3584 OR MYC13E12.33C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37Rv;
 RX MEDLINE-98295987; PubMed-96534230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jorgels K., Krogh A., McLean J., Moule J., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.,
 RT "Complete genome sequence of *Mycobacterium tuberculosis* from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khoulti R., Gill J., Mikula A.,
 RA White W.,
 RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0089 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z95390; CAB08703.1;
 DR EMBL: AE007161; AAK47943.1; ALT_INIT.
 DR TIGR: M3584;
 DR Tuberculist: RV3480C;
 DR InterPro: IPR004255; UPF0089.
 DR Pfam: PF03007; UPF0089; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 427 427 L -> V (IN REF. 2).
 SQ SEQUENCE 497 AA; 53314 MW; 90D661E32B565862 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 497;
 Best Local Similarity 45.5%; Pred. No. 8;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SRXHXSMEX 11
 Db 108 SRLHSHDME 118

```

RESULT 6
RN YNH4_YEAST
RP YNH4_YEAST STANDARD; PRT: 452 AA.
ID P32047;
AC 097107;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 49.2 kDa protein IMP4-MSK1 intergenic region.
GN YN074C OR N2359 OR YMK1.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN-S288c / FY1679;
RX MEDLINE=96267764; PubMed=8701611;
RA Poehlmann R., Philippson P.;
RT "Sequencing a cosmid clone of Saccharomyces cerevisiae chromosome XIV
RT reveals 12 new open reading frames (ORFs) and an ancient duplication
RT of six ORFs."
RL Yeast 12:391-402(1996).

[2]
RN SEQUENCE OF 1-217 FROM N.A.
RP MEDLINE=91202535; PubMed=2016746;
RA Gatti D., Tzagoloff A.;
RT "Structure and evolution of a group of related aminoacyl-tRNA
RT synthetases."
RL J. Mol. Biol. 218:557-568(1991).

[3]
RN -1 SIMILARITY: TO YEAST YIL135C.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X86470; CA60186.1; -;
DR EMBL; Z71350; CA95948.1; -;
DR EMBL; X57360; -; NOT_ANNOTATED_CDS.
DR PIR; S27290; S27290.
DR SGD; S0005018; YNL074C.
KW Hypothetical protein.
FT DOMAIN 211 214 POLY-THR.
FT DOMAIN 271 274 POLY-SER.
SQ SEQUENCE 452 AA; 49210 MW; 9C4FFB02D338D6963E CRC64;

Query Match 68.1%; Score 32; DB 1; Length 452;
Best Local Similarity 54.5%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 SRXHXSMEXR 11
DB 189 SRSHSAAATR 199

```

```

RESULT 7
RN VP41_BPAPS
RP VP41_BPAPS STANDARD; PRT: 460 AA.
ID 097107;
AC 097107;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative protein P41.
GN 41.
OS Bacteriophage APSE-1.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC Unclassified Podoviridae.
OX NCBI_TaxID=106199;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=99420383; PubMed=10489345;
RX van der Wilk F., Dullemeers A.M., Verbeek M., van den Heuvel J.F.J.M.;
RT "Isolation and characterization of APSE-1, a bacteriophage infecting
RT the secondary endosymbiont of acyrthosiphon pisum."
RL Virology 262:104-113(1999).

[2]
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN -----
DR EMBL; AF157835; AAF03984.1; -;
KW Hypothetical protein.
SQ SEQUENCE 460 AA; 51920 MW; 9572842E1FFDDE31 CRC64;

Query Match 68.1%; Score 32; DB 1; Length 460;
Best Local Similarity 55.6%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 1 SRXHXSMEXR 9
DB 259 ARAHYHAME 267

RESULT 8
RN MCRA_BCOLI
ID MCRA_BCOLI STANDARD; PRT: 277 AA.
AC P24200;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5-methylcytosine-specific restriction enzyme A (EC 3.1.21.-)
DE (EcoMCRa).
GN MCRA OR RGIA OR B1159.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=92041637; PubMed=1938927;
RX Hlom K.J., Sedgwick S.G.;
RT "Cloning and structural characterization of the mcra locus of
RT Escherichia coli."
RL J. Bacteriol. 173:7368-7373(1991).

[2]
RN SEQUENCE FROM N.A.
RP Ramalingam R., Prasad R., Shivapriya R., Dharmalingam K.;
RT "Molecular cloning and sequencing of mcra locus and identification of
RT mcra protein in Escherichia coli."
RL J. Biosci. 17:217-232(1992).

[3]
RN SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RX Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RX Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RX Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).

[4]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RC MEDLINE=97061202; PubMed=8905232;
RX Ooshima T., Alha H., Baba T., Fujita K., Hayashi K., Honjo A.,
RX Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RX Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,

```

RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.,
 RT "A 718-bp DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Ref. 3:137-155(1996).
 CC -1- FUNCTION: RESTRICTION OF 5-METHYL AND 5-HYDROXYMETHYLCYTOSINES AT
 CC THE SPECIFIC DNA SEQUENCE C(M)CGG.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M76667; AAA68481.1; -
 DR EMBL: Z19104; CAI79520.1; -
 DR EMBL: AE000215; AAC74243.1; -
 DR EMBL: D90750; BAA35995.1; -
 DR PIR: A41424; A41424.
 DR REBASE: 2832; ECORMCR.
 DR Ecogene: BG10573; MCR.
 DR InterPro: IPR002711; HNH.
 DR InterPro: IPR003615; HNH_nuc.
 DR Pfam: PF01844; HNH; 1.
 DR SMART: SM00507; HNHc; 1.
 DR Hydrolase: Endonuclease; Nuclease; Restriction system;
 KM Complete proteome.
 FT COMPLETE 214 A -> R (IN REF. 1).
 FT SEQUENCE 277 AA; 31390 MW; E5E2627DFDEDEC402 CRC64;
 SQ
 Query Match 66.0%; Score 31; DB 1; Length 277;
 Best Local Similarity 50.0%; Pred. No. 11;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 SRXHSMEXR 11
 DB 69 RKHSLSIDER 78
 RESULT 9
 TR2A_HUMAN STANDARD; PRT; 282 AA.
 ID TR2A_HUMAN
 AC Q13595;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transformer-2 protein homolog (TRA-2 alpha).
 GN TRA2A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96392356; PubMed=8799144;
 RA Daubalder B., Amaya-Manzanares F., Mattox W.;
 RT "A human homologue of the *Drosophila* sex determination factor
 RT transformer-2 has conserved splicing regulatory functions.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:9004-9009(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Oterus;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP FUNCTION, SUBCELLULAR LOCATION, AND PHOSPHORYLATION.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=98206475; PubMed=9546399;
 RA Tacke R., Tohyama M., Ogawa S., Manley J.L.;
 RT "Human Tra2 proteins are sequence-specific activators of pre-mRNA

RT splicing.";
 RL Cell 93:139-148(1998).
 CC -1- FUNCTION: Sequence-specific RNA-binding protein which participates
 CC in the control of pre-mRNA splicing.
 CC -1- SUBUNIT: Binds to A3 enhancer proteins SRP75, SRP55, SRP40 and
 CC SRP30.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- PTM: Phosphorylated in the RS domain.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U53209; AAC50658.1; -
 DR EMBL: BC017094; AAH17094.1; -
 DR HSSP: P19339; 2SXL.
 DR MIM: 602718; -
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 1.
 DR SMART: SM00360; RRM; 1.
 DR PROSITE: PS50102; RRM; 1.
 DR PROSITE: PS00030; RRM_RNP_1; 1.
 DR RNA-binding; mRNA splicing; mRNA processing; Alternative splicing;
 KM Nuclear protein; Phosphorylation.
 FT DOMAIN 30 112 ARG/SER-RICH (RS1 DOMAIN).
 FT DOMAIN 119 197 RRM-BINDING (RRM).
 FT DOMAIN 198 225 LINKER.
 FT DOMAIN 226 282 ARG/SER-RICH (RS2 DOMAIN).
 FT VARSPLIC 1 101 MISSING (IN SHORT ISOFORM).
 FT VARSPLIC 214 214 H -> Q (IN SHORT ISOFORM).
 FT VARSPLIC 215 282 MISSING (IN SHORT ISOFORM).
 FT SEQUENCE 282 AA; 32688 MW; EDB5ABE7BEA023FD CRC64;
 SQ
 Query Match 66.0%; Score 31; DB 1; Length 282;
 Best Local Similarity 54.5%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 SRXHSMEXR 11
 DB 71 SRSHSHSHRR 81
 RESULT 10
 HSFL_ARATH STANDARD; PRT; 495 AA.
 ID HSFL_ARATH
 AC P41151; O23615;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Heat shock factor protein 1 (HSF 1) (Heat shock transcription factor
 DE 1) (HSF1).
 GN HSF1 OR AR4G17750 OR D14910C.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95036006; PubMed=7948881;
 RA Huebel A., Schoeffl F.;
 RT "Arabidopsis heat shock factor: isolation and characterization of the
 RT gene and the recombinant protein.";
 RL Plant Mol. Biol. 26:353-362(1994).
 RN [2]


```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL; AL021287; CAA16086.1; -
CC DR EMBL; AE007128; AAK47410.1; ALT_INIT.
CC DR HSSP; 00192; 10MG.
CC DR TIGR; M3081; -
CC DR TubercuList; RV3001c; -
CC DR InterPro; IPR000506; ACh_isomdctse.
CC DR Pfam; PF01450; IIVC; 1.
CC DR TIGRFAMS; TIGR00465; IIVC; 1.
CC KM Oxidoreductase; Branched-chain amino acid biosynthesis; NADP;
CC FW Complete proteome.
CC FT ACT_SITE 105 105 POTENTIAL.
CC SQ SEQUENCE 333 AA; 36090 MW; 49669E3EDACE25EC CRC64;
CC
CC Query Match 63.88; Score 30; DB 1; Length 333;
CC Best Local Similarity 45.58; Pred. NO. 22;
CC Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
CC
CC QY 1 SRXHXSMEXR 11
CC | : | | : |
CC Db 24 SQGHAHSLSLR 34
CC
CC RESULT 13
CC IIVC_CAMJE ID ID IIVC_CAMJE STANDARD; PRT; 340 AA.
CC AC 09PHN5;
CC DT 16-OCT-2001 (Rel. 4.0, Created)
CC DT 16-OCT-2001 (Rel. 4.0, Last sequence update)
CC DT 15-JUN-2002 (Rel. 4.1, Last annotation update)
CC DE Kcwl- $\alpha$ -acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
CC isomerase/reductase) (Alpha-keto-beta-hydroxyacyl reductoisomerase).
CC GN IIVC OR C30632.
CC OS Campylobacter jejuni.
CC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
CC Campylobacter.
CC OX NCBI_TaxID=197;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=NCCTC 11168;
CC RX MEDLINE=20150912; PubMed=10688204;
CC RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
CC Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
CC Jagsam K., Karlyshew A.V., Moule S., Pallen M.J., Penn C.W.,
CC Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
CC RA Whitehead S., Barrall B.G.;
CC RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
CC RT reveals hypervariable sequences.";
CC RL Nature 403:665-668(2000).
CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
CC = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -1- PATHWAY: Valine and isoleucine biosynthesis; second step.
CC -1- SIMILARITY: BELONGS TO THE KETO-ACID REDUCTOISOMERASE FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL; AL139075; CAB75268.1; -
CC DR InterPro; IPR000506; ACh_isomdctse.
CC DR Pfam; PF01450; IIVC; 1.
CC DR TIGRFAMS; TIGR00465; IIVC; 1.
CC KM Oxidoreductase; Branched-chain amino acid biosynthesis; NADP.
CC FW OXidoreductase; Branched-chain amino acid biosynthesis; NADP.

```


Thu May 1 15:54:32 2003

Db 1014 SRLNNHQMEXR 1024

Search completed: May 1, 2003, 14:33:22
Job time : 5.60938 secs

us-09-446-543a-74.rsp

10

10

10

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 6.1875 Seconds

(without alignments)
170.906 Million cell updates/sec

Title: US-09-446-543a-74

Sequence: 1 SRKXHSMEXR 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: PIR73:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	43	91.5	83	2 JCT607
2	35	74.5	451	2 G01227
3	34	72.3	672	2 A65024
4	34	72.3	672	2 H91046
5	34	72.3	672	2 D85891
6	33	70.2	497	2 E70568
7	32	68.1	196	2 D84351
8	32	68.1	383	2 F71550
9	32	68.1	452	2 S53906
10	32	68.1	648	2 G72279
11	31	66.0	277	2 A41424
12	31	66.0	495	1 S52641
13	30	63.8	330	2 H84075
14	30	63.8	332	1 B70015
15	30	63.8	333	1 JCS166
16	30	63.8	333	1 D70855
17	30	63.8	340	2 G81411
18	30	63.8	393	2 T28994
19	30	63.8	509	2 T36710
20	30	63.8	599	2 F82291
21	30	63.8	1224	2 S70100
22	29	61.7	122	2 G81775
23	29	61.7	122	2 G81775
24	29	61.7	153	2 S33363
25	29	61.7	194	2 A23277
26	29	61.7	292	2 T05317
27	29	61.7	307	2 T17104
28	29	61.7	462	1 Q0BEDA
29	29	61.7	467	2 H88493

ALIGNMENTS

RESULT 1

JCT607
prolactin-releasing peptide - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001

C:Accession: JCT607

R:Yamada, M.; Ozawa, A.; Ishii, S.; Shibusawa, N.; Hashida, T.; Ishizuka, T.; Hosoya, Blochem. Biophys. Res. Commun. 281, 53-56, 2001

A:Title: Isolation and characterization of the rat prolactin-releasing peptide gene:

A:Contents: Spleen

A:Accession: JCT607

A:Molecule type: DNA

A:Residues: 1-83 <YAM>

A:Cross-references: DDBJ:AB040612; DDBJ:AB040613

C:Comment: This peptide induces arachidonic acid metabolite release from rat anterior release, and stimulation of ACTH secretion from the pituitary.

C:Genetics:

A:Gene: PrRP

A:Introns: 33/1

Query Match
Best Local Similarity 91.5%; Score 43; DB 2; Length 83;
Matches 8; Conservative 72.7%; Pred. No. 0.037;
Mismatches 3; Indels 0; Gaps 0;

OY 1 SRKXHSMEXR 11
DB 22 SRKXHSMEXR 32

RESULT 2

G01227

receptin - human

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999

C:Accession: G01227

R:Chao, L.
submitted to the EMBL Data Library, November 1993

A:Reference number: G06292

A:Accession: G01227

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-451 <CHA>

A:Cross-references: EMBL:U03644; NID:9476104; PIDN:AA17653.1; PID:9476105

C:Genetics:

A:Gene: receptin

Query Match
Best Local Similarity 74.5%; Score 35; DB 2; Length 451;
Matches 7; Conservative 63.6%; Pred. No. 9.6;
Mismatches 0; Indels 4; Gaps 0;

OY 1 SRKXHSMEXR 11

DB 360 SRHXHSMEXR 370

RESULT 3

A:Accession: A65024

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: A65024; D49749

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 207, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9276503

A:Accession: A65024

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-672 <BLAT>

A:Cross-references: GB:A6000335; GB:U00096; NID:91788821; PIDN:AMC75535.1; PID:91788827;

A:Experimental source: strain K-12, substrain MG1655

R:Andrews, S.C.; Harrison, P.M.; Guest, J.R.

J. Gen. Microbiol. 137, 361-367, 1991

A:Title: A molecular analysis of the 53.3 minute region of the Escherichia coli linkage

A:Reference number: A49749; MUID:91202105; PMID:2016588

A:Accession: D49749

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-34 <AND>

A:Cross-references: GB:M37689

C:Superfamily: formate hydrogenlyase chain 3

C:Keywords: oxidoreductase

Query Match

Best Local Similarity 72.3%; Score 34; DB 2; Length 672;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Query Match

Best Local Similarity 66.7%; Pred. No. 23;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Query Match

Best Local Similarity 72.3%; Score 34; DB 2; Length 672;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 364 SRHXHSMEXR 372

RESULT 5

D85891

hydrogenase 4 membrane subunit [Imported] - Escherichia coli (strain O157:H7, substra

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: D85891

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamocasts, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: D85891

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-672 <STO>

A:Cross-references: GB:A600174; NID:912516867; PIDN:AA657592.1; GSPDB:GN00145; UMGF:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: hlyE

C:Superfamily: formate hydrogenlyase chain 3

Query Match

Best Local Similarity 72.3%; Score 34; DB 2; Length 672;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Query Match

Best Local Similarity 66.7%; Pred. No. 23;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Query Match

Best Local Similarity 70.2%; Score 33; DB 2; Length 497;

Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Query Match

Best Local Similarity 45.5%; Pred. No. 28;

Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 108 SRHXHSMEXR 118

RESULT 7

D84351

hypothetical protein Vng2006 [Imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: D84351

R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483; PMID:11016950

R:Ramalingam, R.; Prasad, R.; Shivapriya, R.; Dharmalingam, K.
 J. Biosci. 17, 217-232, 1992
 A:Title: Molecular cloning and sequencing of mcrA locus and identification of McrA prote
 A:Reference number: 548131
 A:Accession: 548131
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-277 <RAN>
 A:Cross-references: EMBL:219104; NID:941984; PIDN:CA479520.1; PID:941985
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Petina, N.T.; Burland, V.; Riley, M.; C
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D64861
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-277 <BALT>
 A:Cross-references: GB:AEO00215; GB:U00096; NID:q1787405; PIDN:ANC74243.1; PID:q1787406;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: mcrA; mcrI; rqa
 A:Map position: 25 min
 C:Superfamily: Escherichia coli 5-methylcytosine-specific restriction enzyme A
 C:Keywords: endonuclease; hydrolase; restriction modification system

Query Match 66.0%; Score 31; DB 2; Length 277;
 Best Local Similarity 50.0%; Pred. No. 39;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 RXHXSMEXR 11
 Db 69 RKHMHSLEDR 78

RESULT 12

S52641
 heat shock transcription factor HSF1 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Variety: Columbia
 C:Date: 01-Aug-1995 #sequence_revision 24-Mar-1999 #text_change 16-Jun-2000
 A:Accession: F71447; S52641; S62227; S38873
 R:Bayan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Gled
 avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
 erhoft, A.; Moores, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
 C.; Chaiwalit, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
 A:Reference number: A71400; MUID:98121113; PMID:9461215
 A:Accession: F71447
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-495 <BRV>
 A:Cross-references: GB:297344; NID:92245126; PIDN:CA10555.1; PID:92245134
 R:Huebel, A.; Schoeffl, F.
 Plant Mol. Biol. 26, 353-362, 1994
 A:Title: Arabidopsis heat shock factor: isolation and characterization of the gene and t
 A:Reference number: S52641; MUID:95036006; PMID:7948881
 A:Accession: S52641
 A:Molecule type: DNA
 A:Residues: 1-215, 'M', 217-284, 'D', 286-335, 'T', 337-389, 'Y', 391-469, 474-495 <HUE>
 A:Cross-references: EMBL:X76167
 A:Accession: S62227
 A:Molecule type: mRNA
 A:Residues: 31-215, 'M', 217-284, 'D', 286-335, 'T', 337-389, 'Y', 391-469, 474-495 <HUE>
 R:Huebel, A.; Schoeffl, F.
 Submitted to the EMBL Data Library, November 1993
 A:Description: Arabidopsis heat shock factor: isolation and characterization of the gene
 A:Reference number: S38873
 A:Accession: S38873
 A:Molecule type: DNA

A:Residues: 1-215, 'M', 217-284, 'D', 286-335, 'T', 337-389, 'Y', 391-461, 474-495 <HUE>
 A:Cross-references: EMBL:X76167; NID:9429154; PIDN:CA53761.1; PID:9429155
 C:Genetics:
 A:Gene: HSF1
 A:Map position: 4COP9-4G3845
 A:Introns: 111/3
 A:Function: <TEC>
 A:Description: transcription factor that binds to heat shock promoter elements
 C:Superfamily: tomato heat shock transcription factor HSF8; HSF DNA-binding domain ho
 F;35-148/Domain: HSF DNA-binding domain homology <HSE>
 F;181-202/Region: leucine zipper

Query Match 66.0%; Score 31; DB 1; Length 495;
 Best Local Similarity 55.6%; Pred. No. 71;
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMSE 9
 Db 277 SNGSHSLE 285

RESULT 13

H84075
 thioresdoxin reductase BH3408 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 03-Aug-2001
 A:Accession: H84075
 R:Takaki, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: H84075
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-330 <STO>
 A:Cross-references: GB:AP001518; GB:BA000004; NID:q10175792; PIDN:BA07127.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3408
 C:Superfamily: thioresdoxin reductase; thioresdoxin reductase homology

Query Match 63.8%; Score 30; DB 2; Length 330;
 Best Local Similarity 62.5%; Pred. No. 75;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 RXHXSMSE 9
 Db 190 RAHEHSVE 197

RESULT 14

B70015
 thioresdoxin reductase homolog yumC - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 A:Accession: B70015
 R:Funst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlingsson, J.; Fabbri, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Hensut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koertel, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y., M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scani
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akuch, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tozoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.R.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: B70015
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-332 <KUN>
 A/Cross-references: GB:299120; GB:AL009126; NID:g2635613; PIDN:CAB15201.1; PTD:g2635708
 A/Experimental source: strain 168
 C/Genetics:
 A/Gene: yunc
 C/Superfamily: thioresoxin reductase; thioresoxin reductase homology

Query Match 63.8%; Score 30; DB 1; Length 332;
 Best Local Similarity 62.5%; Pred. No. 76;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 RXHXSM 9
 | | | | |
 Db 190 RAHEHYE 197

RESULT 15

JC5166
 ketol-acid reductoisomerase (EC 1.1.1.86) - Mycobacterium avium
 N/Alternate names: acetoxyhydroxy acid isomerase
 C/Species: Mycobacterium avium
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: JC5166
 R/Gusberti, L.; Cantoni, R.; De Rossi, E.; Branzoni, M.; Riccardi, G.
 Gene 177, 83-85, 1996
 A/Title: Cloning and sequencing of the llyNC gene cluster from Mycobacterium avium.
 A/Reference number: JC5164; MID:97080504; PMID:8921849
 A/Accession: JC5166
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-333 <GUS>
 A/Cross-references: GB:L49392; NID:g1196506; PIDN:AB38428.1; PID:g1196509
 C/Genetics:
 A/Gene: llyC
 C/Superfamily: Methanococcus ketol-acid reductoisomerase; ketol-acid reductoisomerase hc
 C/Keywords: Intramolecular transferase; isoleucine-valine biosynthesis; isomerase; oxid
 F;18-200/Domain: ketol-acid reductoisomerase homology <KAR>

Query Match 63.8%; Score 30; DB 1; Length 333;
 Best Local Similarity 45.5%; Pred. No. 76;
 Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
 | | | | |
 Db 24 SOGHARSLR 34

Search completed: May 1, 2003, 14:31:42
 Job time: 10.1875 secs

22

22

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:33:06 ; Search time 6.875 Seconds

(without alignments)
138.059 Million cell updates/sec

Title: US-09-446-543a-74

Perfect score: 47

Sequence: 1 SRXHXSMEXR 11

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	91.5	31	10	US-09-932-161-14
2	43	91.5	31	12	US-10-044-592-4
3	43	91.5	31	12	US-10-044-592-5
4	43	91.5	70	12	US-10-044-592-90
5	43	91.5	86	12	US-10-044-592-1
6	43	91.5	86	12	US-10-044-592-96
7	43	91.5	91	12	US-10-044-592-94
8	43	89.4	29	12	US-10-044-592-26
9	42	89.4	31	10	US-09-932-161-13
10	42	89.4	31	10	US-09-932-161-15
11	42	89.4	31	12	US-10-044-592-39
12	42	89.4	32	12	US-10-044-592-40
13	42	89.4	33	12	US-10-044-592-41
14	42	89.4	40	12	US-10-044-592-80
15	42	89.4	87	12	US-10-044-592-82
16	42	89.4	98	12	US-10-044-592-28
17	42	89.4	98	12	US-10-044-592-38
18	42	89.4	98	12	US-10-044-592-82
19	42	89.4	98	12	US-10-044-592-84

20	42	89.4	98	12	US-10-044-592-86	Sequence 86, App1
21	42	89.4	98	12	US-10-044-592-88	Sequence 88, App1
22	38	80.9	9	12	US-10-044-592-7	Sequence 7, App11
23	33	70.2	25	12	US-10-044-592-78	Sequence 78, App11
24	32	68.1	383	10	US-09-841-132-580	Sequence 580, App
25	31	66.0	103	10	US-09-841-132-580	Sequence 1964, Ap
26	30	63.8	236	9	US-09-984-245-318	Sequence 318, App
27	30	63.8	236	9	US-09-984-245-318	Sequence 318, App
28	30	63.8	236	9	US-09-984-245-318	Sequence 318, App
29	30	63.8	236	9	US-09-984-245-318	Sequence 318, App
30	30	63.8	236	9	US-10-143-090-318	Sequence 318, App
31	30	63.8	1224	10	US-09-801-368-222	Sequence 222, App
32	29	61.7	18	10	US-09-864-761-37659	Sequence 37659, A
33	29	61.7	525	10	US-09-804-551B-36	Sequence 36, App1
34	29	61.7	580	10	US-09-925-301-943	Sequence 943, App
35	29	61.7	774	10	US-09-974-298-122	Sequence 122, App
36	29	61.7	774	10	US-09-782-080-16	Sequence 16, App1
37	28	59.6	774	10	US-09-909-743-7	Sequence 7, App11
38	28	59.6	44	10	US-09-864-761-37970	Sequence 37970, A
39	28	59.6	113	10	US-09-916-790-27	Sequence 27, App1
40	28	59.6	338	9	US-09-738-626-6959	Sequence 6959, Ap
41	28	59.6	390	10	US-09-901-151-5	Sequence 5, App11
42	28	59.6	397	10	US-09-901-151-6	Sequence 6, App11
43	28	59.6	414	10	US-09-901-151-2	Sequence 2, App11
44	28	59.6	575	9	US-10-014-101-2	Sequence 4, App11
45	28	59.6	577	10	US-09-815-242-10193	Sequence 2, App11

ALIGNMENTS

RESULT 1
US-09-932-161-14
; Sequence 14, Application US/09932161
; Patent No. US20020037553A1
GENERAL INFORMATION:
; APPLICANT: Civeilli, Olivier
; TITLE OF INVENTION: Screening and Therapeutic Methods For
; FILE REFERENCE: P-UC 4679
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US/09/932,161
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Rattus
US-09-932-161-14
Query Match 91.5%; Score 43; DB 10; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.049;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
DB 1 SRXHXSMEXR 11
1 SRXHXSMEXR 11
US-10-044-592-4
; Sequence 4, Application US/10044592
; Patent No. US20020143152A1
GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10

PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 4
LENGTH: 31
TYPE: PRT
ORGANISM: Murine
US-10-044-592-4

Query Match 91.5%; Score 43; DB 12; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.049;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
|||
Db 1 SRAHQSMETR 11

RESULT 3
US-10-044-592-5
Sequence 5, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 5
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(31)
OTHER INFORMATION: antigen
US-10-044-592-5

Query Match 91.5%; Score 43; DB 12; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.049;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
|||
Db 1 SRAHQSMETR 11

RESULT 4
US-10-044-592-90
Sequence 90, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 9
LENGTH: 70
TYPE: PRT
ORGANISM: Rat
US-10-044-592-90

PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 90
LENGTH: 70
TYPE: PRT
ORGANISM: Rat
US-10-044-592-90

Query Match 91.5%; Score 43; DB 12; Length 70;
Best Local Similarity 72.7%; Pred. No. 0.11;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 14
|||
Db 22 SRAHQSMETR 32

RESULT 5
US-10-044-592-1
Sequence 1, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 1
LENGTH: 82
TYPE: PRT
ORGANISM: Murine
US-10-044-592-1

Query Match 91.5%; Score 43; DB 12; Length 82;
Best Local Similarity 72.7%; Pred. No. 0.13;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
|||
Db 21 SRAHQSMETR 31

RESULT 6
US-10-044-592-96
Sequence 96, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 96
LENGTH: 31
TYPE: PRT
ORGANISM: Murine
US-10-044-592-96

NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 96
; LENGTH: 86
; TYPE: PRT
; ORGANISM: mammalian
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1224)..(1243)
; OTHER INFORMATION: Bracket region depicted in FIG 39.
US-10-044-592-96

Query Match 91.5%; Score 43; DB 12; Length 86;
Best Local Similarity 72.7%; Pred. No. 0.14;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 22 SRAHQHSMETR 32

RESULT 7
US-10-044-592-94
; Sequence 94, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 94
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Mammalian
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(31)
; OTHER INFORMATION: primer
; NAME/KEY: misc_feature
; LOCATION: (925)..(955)
; OTHER INFORMATION: primer
; NAME/KEY: misc_feature
; LOCATION: (1)..(955)
; OTHER INFORMATION: Insert fragment of pmgB3 as depicted in FIG 36 and 37.
US-10-044-592-94

Query Match 91.5%; Score 43; DB 12; Length 91;
Best Local Similarity 72.7%; Pred. No. 0.15;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 22 SRAHQHSMETR 32

RESULT 8
US-10-044-592-26
; Sequence 26, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use

FILE REFERENCE: 2463052P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 26
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-26

Query Match 89.4%; Score 42; DB 12; Length 29;
Best Local Similarity 72.7%; Pred. No. 0.071;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 1 SRAHQHSMETR 11

RESULT 9
US-09-932-161-13
; Sequence 13, Application US/09932161
; Patent No. US2002037533A1
; GENERAL INFORMATION:
; APPLICANT: Clivelli, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Screening and Therapeutic Methods for
; TITLE OF INVENTION: Promoting Wakefulness and Sleep
; FILE REFERENCE: P-UC 4679
; CURRENT APPLICATION NUMBER: US/09/932,161
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/560,915
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-932-161-13

Query Match 89.4%; Score 42; DB 10; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
Db 1 SRAHQHSMETR 11

RESULT 10
US-09-932-161-15
; Sequence 15, Application US/09932161
; Patent No. US2002037533A1
; GENERAL INFORMATION:
; APPLICANT: Clivelli, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Screening and Therapeutic Methods for
; TITLE OF INVENTION: Promoting Wakefulness and Sleep
; FILE REFERENCE: P-UC 4679
; CURRENT APPLICATION NUMBER: US/09/932,161
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/560,915
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 15
LENGTH: 31
TYPE: PRT
ORGANISM: Homo Sapien
us-09-932-161-15

Query Match 89.4%; Score 42; DB 10; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.076;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 1 SRAHQHSMETR 11

RESULT 11
US-10-044-592-39

Sequence 39, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji

TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P

CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10

PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10

PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27

PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28

NUMBER OF SEQ ID NOS: 96
SOFTWARE:

SEQ ID NO 39
LENGTH: 31
TYPE: PRT

ORGANISM: Bovine
US-10-044-592-39

Query Match 89.4%; Score 42; DB 12; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.076;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 1 SRAHQHSMETR 11

RESULT 12
US-10-044-592-40

Sequence 40, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji

TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P

CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10

PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10

PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27

PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28

NUMBER OF SEQ ID NOS: 96
SOFTWARE:

SEQ ID NO 40
LENGTH: 32
TYPE: PRT

ORGANISM: Bovine
US-10-044-592-40

Query Match 89.4%; Score 42; DB 12; Length 32;
Best Local Similarity 72.7%; Pred. No. 0.079;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 1 SRAHQHSMETR 11

RESULT 13
US-10-044-592-41

Sequence 41, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji

TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P

CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10

PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10

PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27

PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28

NUMBER OF SEQ ID NOS: 96
SOFTWARE:

SEQ ID NO 41
LENGTH: 33
TYPE: PRT

ORGANISM: Bovine
US-10-044-592-41

Query Match 89.4%; Score 42; DB 12; Length 33;
Best Local Similarity 72.7%; Pred. No. 0.082;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
DB 1 SRAHQHSMETR 11

RESULT 14
US-10-044-592-80

Sequence 80, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji

TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P

CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10

PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10

PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27

PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28

NUMBER OF SEQ ID NOS: 96
SOFTWARE:

SEQ ID NO 80
LENGTH: 40
TYPE: PRT

ORGANISM: Bovine
US-10-044-592-80

NAME/KEY: misc.feature
LOCATION: (99)-(126)
OTHER INFORMATION: primer

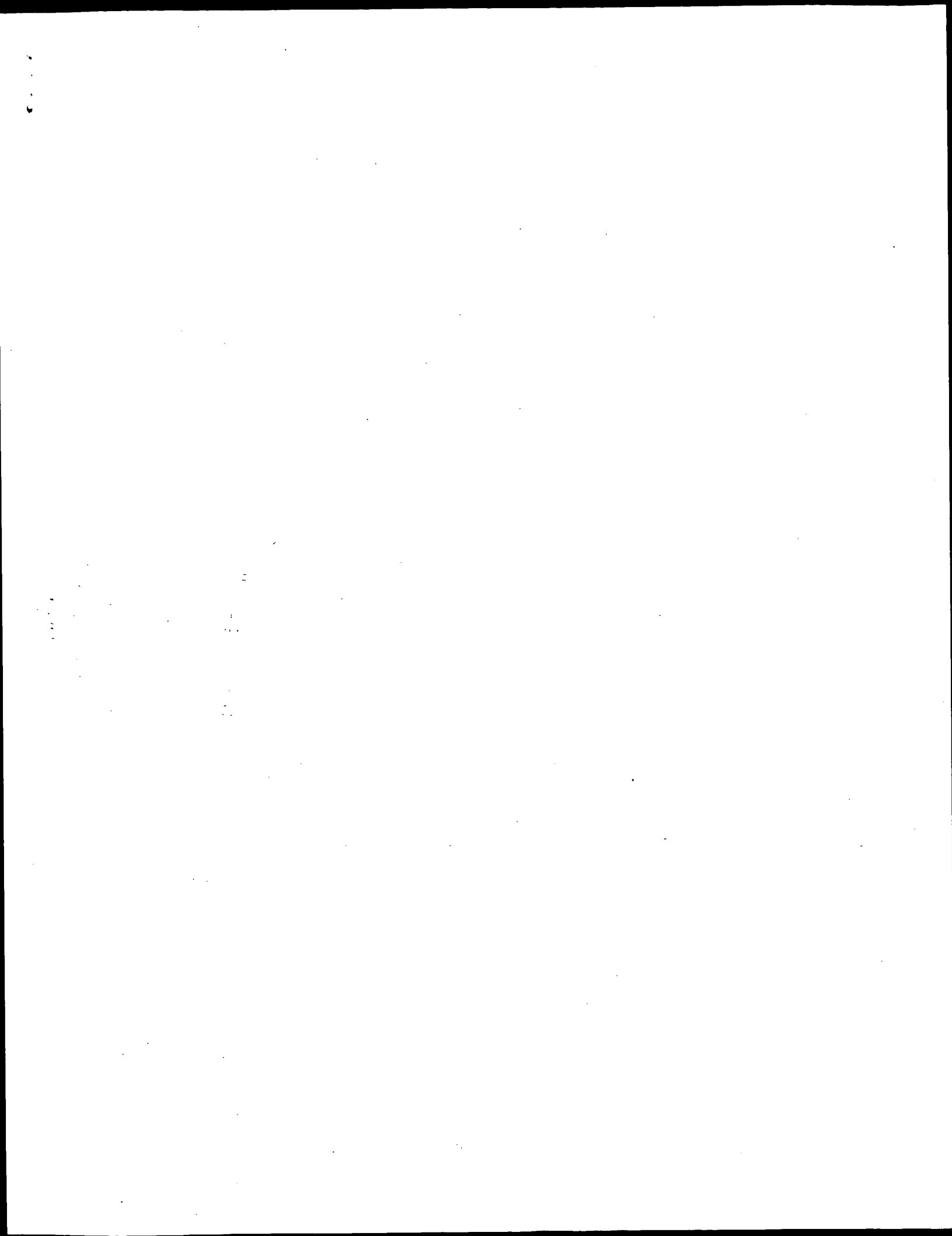
Query Match 89.4%; Score 42; DB 12; Length 40;
Best Local Similarity 72.7%; Pred. No. 0.099;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 SRXHXSMEXR 11
||| |||||
Db 23 SRAHSHMEIR 33

RESULT 15
US-10-044-592-92
; Sequence 92, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 92
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-044-592-92

Query Match 89.4%; Score 42; DB 12; Length 87;
Best Local Similarity 72.7%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
||| |||||
Db 23 SRAHSHMEIR 33

Search completed: May 1, 2003, 14:46:07
Job time : 7.875 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:13 ; Search time 5.32812 Seconds

(Without alignments)
60.744 Million cell updates/sec

Title: US-09-446-543a-74

Sequence: 1 SRXHXSMEXR 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/1/1aa/5a COMB pep:*
2: /cgn2_6/prodata/1/1aa/5b COMB pep:*
3: /cgn2_6/prodata/1/1aa/6a COMB pep:*
4: /cgn2_6/prodata/1/1aa/6b COMB pep:*
5: /cgn2_6/prodata/1/1aa/6c COMB pep:*
6: /cgn2_6/prodata/1/1aa/6d COMB pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	91.5	31	3	US-09-105-678A-8
2	43	91.5	31	3	US-09-105-678A-37
3	43	91.5	31	4	US-09-172-353-4
4	43	91.5	31	4	US-08-776-971-47
5	43	91.5	31	4	US-09-421-208-8
6	43	91.5	31	4	US-09-421-208-37
7	43	91.5	31	4	US-09-560-915-14
8	43	91.5	32	3	US-09-105-678A-38
9	43	91.5	32	3	US-08-776-971-48
10	43	91.5	32	4	US-09-421-208-38
11	43	91.5	33	3	US-09-105-678A-39
12	43	91.5	33	4	US-08-776-971-49
13	43	91.5	33	4	US-09-421-208-39
14	43	91.5	83	4	US-08-776-971-45
15	43	91.5	83	4	US-08-776-971-124
16	43	91.5	83	4	US-08-776-971-137
17	42	89.4	15	4	US-08-776-971-92
18	42	89.4	29	3	US-09-105-678A-29
19	42	89.4	29	4	US-08-776-971-3
20	42	89.4	29	4	US-09-421-208-29
21	42	89.4	31	3	US-09-105-678A-7
22	42	89.4	31	3	US-09-105-678A-9
23	42	89.4	31	3	US-09-105-678A-31
24	42	89.4	31	3	US-09-105-678A-43
25	42	89.4	31	4	US-08-776-971-5
26	42	89.4	31	4	US-08-776-971-61
27	42	89.4	31	4	US-08-776-971-97

28	42	89.4	31	4	US-09-421-208-7	Sequence 7, Appl
29	42	89.4	31	4	US-09-421-208-9	Sequence 9, Appl
30	42	89.4	31	4	US-09-421-208-31	Sequence 31, Appl
31	42	89.4	31	4	US-09-421-208-43	Sequence 43, Appl
32	42	89.4	31	4	US-09-560-915-15	Sequence 15, Appl
33	42	89.4	31	4	US-09-105-678A-32	Sequence 32, Appl
34	42	89.4	32	3	US-09-105-678A-44	Sequence 44, Appl
35	42	89.4	32	4	US-08-776-971-6	Sequence 6, Appl
36	42	89.4	32	4	US-09-421-208-32	Sequence 32, Appl
37	42	89.4	32	4	US-09-421-208-32	Sequence 32, Appl
38	42	89.4	32	4	US-09-421-208-32	Sequence 32, Appl
39	42	89.4	33	3	US-09-105-678A-43	Sequence 43, Appl
40	42	89.4	33	3	US-09-105-678A-45	Sequence 45, Appl
41	42	89.4	33	4	US-08-776-971-7	Sequence 7, Appl
42	42	89.4	33	4	US-08-776-971-63	Sequence 63, Appl
43	42	89.4	33	4	US-09-421-208-33	Sequence 33, Appl
44	42	89.4	33	4	US-09-421-208-33	Sequence 33, Appl
45	42	89.4	33	4	US-09-421-208-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-09-105-678A-8

Sequence 8, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-8

Query Match 91.5% Score 43; DB 3; Length 31;
Best Local Similarity 72.7% Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11

Db 1 SRHXHSMETR 11

RESULT 2

US-09-105-678A-37
Sequence 37, Application US/09105678A
Patent No. 6103862

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION NUMBER: JP 172118/1997

ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-37

Query Match

Best Local Similarity 91.5%; Score 43; DB 3; Length 31;

Pred. No. 0.013;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRHXHSMETR 11

Db 1 SRHXHSMETR 11

RESULT 3

US-09-172-353-4

Sequence 4, Application US/09172353

Patent No. 6197530

GENERAL INFORMATION:

APPLICANT: Stricker-Kongra, Alain

APPLICANT: Gu, Wei

TITLE OF INVENTION: GPR10 AS A TARGET FOR IDENTIFYING WEIGHT MODULATING COMPOUNDS

FILE REFERENCE: 07334/102001

CURRENT APPLICATION NUMBER: US/09/172,353

CURRENT FILING DATE: 1998-10-14

NUMBER OF SEQ ID NOS: 7

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 31

TYPE: PRT
ORGANISM: Mus musculus
US-09-172-353-4

Query Match

Best Local Similarity 91.5%; Score 43; DB 4; Length 31;

Pred. No. 0.013;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRHXHSMETR 11

Db 1 SRHXHSMETR 11

RESULT 4

US-08-776-971-47

Sequence 47, Application US/08776971B

Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Hanuma, Shuji

APPLICANT: Habata, Yugo

APPLICANT: Kawamata, Yuji

APPLICANT: Hosoya, Masaki

APPLICANT: Fujii, Ryo

APPLICANT: Fukusumi, Shoji

APPLICANT: Kitada, Chieko

TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE

NUMBER OF SEQUENCES: 140

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-08-776-971-47

Query Match

91.5%; Score 43; DB 4; Length 31;

Best Local Similarity 72.7%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
||| |||||
DB 1 SRAHQHSMETR 11

RESULT 5

US-09-421-208-8
Sequence 8, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ. ID NO.: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-8

Query Match 91.5%; Score 43; DB 4; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
||| |||||
DB 1 SRAHQHSMETR 11

RESULT 6

US-09-421-208-37
Sequence 37, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ. ID NO.: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-37

Query Match 91.5%; Score 43; DB 4; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
||| |||||
DB 1 SRAHQHSMETR 11

RESULT 7

US-09-560-915-14
Sequence 14, Application US/09560915
Patent No. 6383764
GENERAL INFORMATION:
APPLICANT: Clavelli, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Therapeutic Compositions and Methods
TITLE OF INVENTION: Relating to Prolactin Releasing Peptide (PRP)
FILE REFERENCE: P-00 3534
CURRENT APPLICATION NUMBER: US/09/560,915
CURRENT FILING DATE: 2000-04-28
NUMBER OF SEQ. ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ. ID NO. 14
LENGTH: 31
TYPE: PRP
ORGANISM: Rattus
US-09-560-915-14

Query Match 91.5%; Score 43; DB 4; Length 31;
Best Local Similarity 72.7%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRXHXSMEXR 11
||| |||||

DB 1 SRAHSHMETR 11

RESULT 8

US-09-105-678A-38
Sequence 38, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MACOUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A

FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-105-678A-38

Query Match 91.5%; Score 43; DB 3; Length 32;
Best Local Similarity 72.7%; Pred. No. 0.014;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMEXR 11

DB 1 SRAHSHMETR 11

RESULT 9

US-08-776-971-48
Sequence 48, Application US/08776971B
Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Hiruma, Shuji
APPLICANT: Hadata, Yugo
APPLICANT: Kawamata, Yuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Fukusumi, Shoji
APPLICANT: Kitada, ChiekoTITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street
CITY: Boston
STATE: MACOUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 48:

US-08-776-971-48

Query Match 91.5%; Score 43; DB 4; Length 32;
Best Local Similarity 72.7%; Pred. No. 0.014;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SRHXHSMEXR 11

DB 1 SRAHSHMETR 11

RESULT 10

US-09-421-208-38
Sequence 38, Application US/09421208
Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MACOUNTRY: USA
ZIP: 02109COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-38

Query Match 91.5%; Score 43; DB 4; Length 32;
Best Local Similarity 72.7%; Pred. No. 0.014;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
||| |||||
Db 1 SRAHSHMETR 11

RESULT 11
US-09-105-678A-39
Sequence 39, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:

LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-39

Query Match 91.5%; Score 43; DB 3; Length 33;
Best Local Similarity 72.7%; Pred. No. 0.014;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRXHXSMEXR 11
||| |||||
Db 1 SRAHSHMETR 11

RESULT 12
US-08-776-971-49
Sequence 49, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Habata, Yugo
APPLICANT: Kawamata, Yuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Fukusumi, Shoji
APPLICANT: Kitada, Chieko
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 49:

Kitada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 124:
SEQUENCE CHARACTERISTICS:
LENGTH: 83 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 124:
US-08-776-971-124
Query Match 91.5%; Score 43; DB 4; Length 83;
Best Local Similarity 72.7%; P-Val. No. 0.036;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 SRXHXSMEXR 11
||| |||||
Db 22 SRAHSHSMETR 32

Search completed: May 1, 2003, 14:33:54
Job time: 5.32812 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 24.0625 Seconds

(without alignments)
188.386 Million cell updates/sec

Title: US-09-446-543a-73

Perfect score: 109

Sequence: 1 TPDINPAMYXXRGIRPVGRFX 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_RVIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	95.4	98	6	Q8WN12
2	82	75.2	117	13	Q9W624
3	49.5	45.4	664	16	Q9I1W4
4	49	45.0	790	10	Q9M371
5	46.5	42.7	420	2	Q9JL27
6	46	42.2	54	4	Q9UJF9
7	46	42.2	333	16	Q9PH76
8	46	42.2	465	4	Q60687
9	46	42.2	465	4	Q8WM85
10	46	42.2	540	10	Q9L6Z0
11	45	41.3	767	5	Q20170
12	44	40.4	145	5	Q95Z85
13	44	40.4	250	16	Q9I2F2
14	44	40.4	284	17	Q50128
15	44	40.4	284	17	Q8U112
16	44	40.4	419	4	Q9Y276

17	44	40.4	884	15	Q8Q718	Q8Q718 human immun
18	43.5	39.9	1501	10	Q9SD86	Q9SD86 arabidopsis
19	43	39.4	220	16	Q9I022	Q9I022 pseudomonas
20	43	39.4	226	16	Q8RR10	Q8RR10 fuscobacteri
21	43	39.4	232	16	Q9RZ21	Q9RZ21 delinococcus
22	43	39.4	309	5	Q17234	Q17234 caenorhabdi
23	43	39.4	333	4	Q96SD4	Q96SD4 homo sapien
24	43	39.4	370	11	Q9C3V7	Q9C3V7 mus musculu
25	43	39.4	418	11	Q9CZP5	Q9CZP5 mus musculu
26	43	39.4	476	16	Q8XA75	Q8XA75 escherichia
27	43	39.4	813	17	Q8ZXH2	Q8ZXH2 pyrobaculum
28	43	39.4	1488	5	Q20294	Q20294 caenorhabdi
29	42.5	39.0	333	16	Q9RJ10	Q9RJ10 streptomyce
30	42.5	39.0	360	5	Q19879	Q19879 caenorhabdi
31	42	38.5	293	2	Q53855	Q53855 spiroplasma
32	42	38.5	416	16	Q99Z49	Q99Z49 streptococ
33	42	38.5	428	16	Q9PEK7	Q9PEK7 chlamydia m
34	42	38.5	433	16	Q9A382	Q9A382 caulobacter
35	42	38.5	467	11	Q8R054	Q8R054 mus musculu
36	42	38.5	549	2	Q9EW96	Q9EW96 streptomyce
37	41.5	38.1	151	12	Q9IR01	Q9IR01 dashen mos
38	41	37.6	97	2	Q33440	Q33440 pseudomonas
39	41	37.6	128	16	P74747	P74747 synechocyst
40	41	37.6	226	16	Q9A359	Q9A359 caulobacter
41	41	37.6	263	16	Q8XRE9	Q8XRE9 ralsstonia s
42	41	37.6	273	17	Q8ZTH7	Q8ZTH7 pyrobaculum
43	41	37.6	280	4	Q969Z5	Q969Z5 homo sapien
44	41	37.6	284	17	Q9V003	Q9V003 pyrococcus
45	41	37.6	328	4	Q9NV85	Q9NV85 homo sapien

ALIGNMENTS

RESULT 1						
ID	Q8WN12	PRELIMINARY;	PRT;	98 AA.		
AC	Q8WN12;					
DT	01-MAR-2002 (TREMBLrel. 20, Created)					
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)					
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)					
DE	Proprolactin-releasing peptide.					
OS	Ovis aries (Sheep).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;					
OC	Bovidae; Caprinae; Ovis.					
OX	NCBI_TaxID=9940;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Curtlewis J.D., Kusters D.H.L., Barclay J.T., Anderson S.T.;					
RT	"Prolactin-releasing peptide (PrRP) in the ewe: cDNA cloning, mRNA					
RT	distribution and effects on prolactin secretion in vitro and in					
RT	vivo."					
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AF450453; AAL47178.1;					
SO	SEQUENCE 98 AA; 10513 MW; 2A53331ED62CA85 CRC64;					
Query Match	95.4%; Score 104; DB 6; Length 98;					
Best Local Similarity	90.0%; Pred. No. 1.2e-09;					
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;					
OY	1 TPDINPAMYXXRGIRPVGRF 20					
DB	34 TPDINPAMYAGRGIRPVGRF 53					
RESULT 2						
ID	Q9W624	PRELIMINARY;	PRT;	117 AA.		
AC	Q9W624;					
DT	01-NOV-1999 (TREMBLrel. 12, Created)					
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)					
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)					

DE C-RF amide.
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Carassius.
 CX NCBI_TaxID=7957;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Satate H., Minakata H., Fujimoto M.,
 RT "Carassius pinnatus (C-RF amide).";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB020024; BAA/6662.1;
 SQ SEQUENCE 117 AA; 12879 MW; D5DC4CB22038C2B0 CRC64;
 Query Match 75.2%; Score 82; DB 13; Length 117;
 Best Local Similarity 60.0%; Pred. No. 6; 3e-06;
 Matches 12; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
 DB 56 SPEIDPFMYVGRGVPRIGRF 75
 RESULT 3
 ID 0911W4 PRELIMINARY; PRT; 664 AA.
 AC 0911W4;
 DT 01-MAR-2001 (TRENBLREL. 16, Created)
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
 DE Hypothetical protein PA2151.
 GN PA2151.
 OS Pseudomonas aeruginosa.
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas
 CX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RL MEDLINE-20437337; PubMed-10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huhagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Slier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL: AE004642; AGO5539.1;
 DR InterPro: IPR001589; Actbind_actin.
 DR InterPro: IPR000461; Alpha_amyrase.
 DR Pfam: PF00128; alpha-amyrase; 1.
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 664 AA; 76329 MW; 8F59FEED54C308AD CRC64;
 Query Match 45.4%; Score 49.5; DB 16; Length 664;
 Best Local Similarity 62.5%; Pred. No. 8.9;
 Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;
 QY 1 TPDINPAMYXXRGIRP 16
 DB 478 TPDINP-WFLQSGRP 492
 RESULT 4
 ID 09M371 PRELIMINARY; PRT; 790 AA.
 AC 09M371;
 DT 01-OCT-2000 (TRENBLREL. 15, Created)
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)

DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
 DE Hypothetical 87.4 kDa protein.
 GN F15G16.60.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA De Haan M., Maarse A.C., Grievell L.A., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Quettier F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL132959; CAB71097.1;
 KW Hypothetical protein.
 SQ SEQUENCE 790 AA; 87376 MW; B222724B75690F30 CRC64;

QY 2 PDINPAMYXXRGIRPVGRF 20
 DB 366 PPHNPRYCGSRGLQPHGRW 384
 Query Match 45.0%; Score 49; DB 10; Length 790;
 Best Local Similarity 47.4%; Pred. No. 13;
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 RESULT 5
 ID 093L27 PRELIMINARY; PRT; 420 AA.
 AC 093L27;
 DT 01-DEC-2001 (TRENBLREL. 19, Created)
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
 DE Chain length factor-like protein.
 GN AUR2B.
 OS Streptomyces aureofaciens.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomycetes.
 CX NCBI_TaxID=1894;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CCM3239;
 RA Kormanec J., Bistakova J., Novakova R., Homerova D., Reznichova B.;
 RT "Cloning and characterization of a new polyketide gene cluster in
 Streptomyces aureofaciens CCM3239.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY033994; AAK61719.1;
 DR InterPro: IPR000794; Ketoacyl-synt.
 DR Pfam: PF00109; ketoacyl-synt; 1.
 DR Pfam: PF02801; ketoacyl-synt_C; 1.
 SQ SEQUENCE 420 AA; 43011 MW; 3C27E22BBE86C2DEA CRC64;
 Query Match 42.7%; Score 46.5; DB 2; Length 420;
 Best Local Similarity 52.6%; Pred. No. 17;
 Matches 10; Conservative 1; Mismatches 3; Indels 5; Gaps 1;
 QY 7 AMYXX-----RGIRPVGRF 20
 DB 37 AMMAAVLRGSGIRPVGRF 55
 RESULT 6
 ID 09UUF9 PRELIMINARY; PRT; 54 AA.
 AC 09UUF9;
 DT 01-MAY-2000 (TRENBLREL. 13, Created)
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
 DT 01-MAY-2000 (TRENBLREL. 13, Last annotation update)
 DE D0479J7.3 (Sushi-repeat protein (SRPUL)) (Fragment).

GN DJ479J7.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lawlor S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035608; CAB55682.1;
 FT NON TER 54 54
 SO SEQUENCE 54 AA; 6110 MW; E2F3C39F7B961A9F CRC64;

Query Match 42.2%; Score 46; DB 4; Length 54;
 Best Local Similarity 43.8%; Pred. No. 2.4;
 Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRP 16
 II : ||| | |
 DB 18 TPVAVPTWYAGSGYYP 33

RESULT 7
 O9PH76 PRELIMINARY; PRT; 333 AA.
 AC O9PH76;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hydroxybenzoate octaprenyltransferase.
 GN XF0068.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 RX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-9A5C;
 RC MEDLINE-20365717; PubMed-10910347;
 RA Simpson A.J.G., Reimach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Coluto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.P., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kurama E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Lemos M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
 RA de Oliveira B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL; AF003860; AAF82881.1;
 DR InterPro; IPR000537; UBLA.
 DR Pfam; PF01040; UBLA.1.
 DR PROSITE; PS00943; UBLA; UNKNOWN_1.
 KW Complete proteome.
 SO SEQUENCE 333 AA; 37931 MW; ECF3F4716C962B95 CRC64;

Query Match 42.2%; Score 46; DB 16; Length 333;
 Best Local Similarity 53.3%; Pred. No. 16;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 4 INPAMYXXRGIRPVG 18
 :: | | | | |
 DB 54 IDPYKRLARGDRPVG 68

RESULT 8
 O60687 PRELIMINARY; PRT; 465 AA.
 AC O60687;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sush1-repeat protein.
 GN SRPUL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kurosawa H., Inukai T., Inaba T., Gai K., Chang K.-S., Sinjo T.,
 RA Rakesstraw K.M., Naeve C.W., Look T.A.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF005567; AAC15765.1;
 DR InterPro; IPR001128; Cytochrome_P450.
 DR InterPro; IPR003410; Hyalin.
 DR InterPro; IPR000436; Sush1_SCR_CCP.
 DR Pfam; PF02494; HYR.1.
 DR Pfam; PF00084; sush1.3.
 DR SMART; SM00032; CCP.3.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 SO SEQUENCE 465 AA; 52971 MW; 4D752B187FE3EEB8 CRC64;

Query Match 42.2%; Score 46; DB 4; Length 465;
 Best Local Similarity 43.8%; Pred. No. 23;
 Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRP 16
 II : ||| | |
 DB 18 TPVAVPTWYAGSGYYP 33

RESULT 9
 O8WW85 PRELIMINARY; PRT; 465 AA.
 AC O8WW85;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sush1-repeat protein.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE-PLACENTA;
 RA Straube R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC020733; AAR20733.1;
 DR InterPro; IPR001128; Cytochrome_P450.
 DR InterPro; IPR003410; Hyalin.
 DR InterPro; IPR000436; Sush1_SCR_CCP.
 DR Pfam; PF02494; HYR.1.
 DR Pfam; PF00084; sush1.3.
 DR SMART; SM00032; CCP.3.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 SO SEQUENCE 465 AA; 52957 MW; 3D7229487DA1B8BD CRC64;

RP SEQUENCE FROM N.A.

DT 01-MAR-2001 (Tremblé, 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)
 DT 01-OCT-2001 (TReMBLrel. 18, last annotation update)
 DE Hypothetical protein PA1952.
 GN PA1952.
 OS Pseudomonas aeruginosa.
 OC Bacteria: Proteobacteria; gamma subdivision: Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickley M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larijs K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RL Nature 406:959-964 (2000).
 DR EMBL: AE004622; AAG05340.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 250 AA; 25619 MW; B997F6BE28D792C2 CRC64;
 Query Match 40.4%; Score 44; DB 16; Length 250;
 Best Local Similarity 47.1%; Pred. No. 26;
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 PDINPAMYXXRGIRPVGR 18
 Db 233 PALNCMEQLRALRPSG 249
 RESULT 14
 ID 050128 PRELIMINARY; PRT; 284 AA.
 AC 050128;
 DT 01-JUN-1998 (TReMBLrel. 06, Created)
 DT 01-JUN-1998 (TReMBLrel. 06, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE Hypothetical protein PH1420.
 GN PH1420.
 OS Pyrococcus horikoshii.
 OC Archaea: Euryarchaeota; Thermococci; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
 RA Aoki K.-I., Yoshitawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
 RL DNA Res. 5:55-76 (1998).
 DR EMBL: AP000006; BAA30526.1;
 DR InterPro: IPR001330; Prenyltrans.
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00432; Prenyltrans; 2.
 DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 284 AA; 32319 MW; 8E0E7BC3711D3815 CRC64;
 Query Match 40.4%; Score 44; DB 17; Length 284;
 Best Local Similarity 44.4%; Pred. No. 30;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 2 PDINPAMYXXRGIRPVGR 19

Db 217 PYIEPTFYALRGLELIGR 234

RESULT 15

ID 080112 PRELIMINARY; PRT; 284 AA.
 AC 080112;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Hypothetical protein pfl1418.
 GN Pfl1418.
 OS Pyrococcus furiosus.
 OC Archaea: Euryarchaeota; Thermococci; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC1 / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Welts R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the Pyrococcus furiosus genome."
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE010244; AAB1342.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 284 AA; 32302 MW; 1D4C5746A1831390 CRC64;
 Query Match 40.4%; Score 44; DB 17; Length 284;
 Best Local Similarity 44.4%; Pred. No. 30;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 2 PDINPAMYXXRGIRPVGR 19
 Db 217 PYIEPTFYALRGLELIGR 234

Search completed: May 1, 2003, 14:32:57
 Job time : 28.0625 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 12.375 Seconds

(Without alignments)
170.906 Million cell updates/sec

Title: US-09-446-543A-73

Perfect score: 109

Sequence: 1 TPDINPAWYXXRGIRPVGRFX 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database : PIR-73:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	104	95.4	83	2	JC7607	prolactin-releasin
2	49.5	45.4	664	2	F8376	conserved hypothet
3	49	45.0	790	2	T47959	hypothetical prote
4	46	42.2	333	2	H82852	hydroxybenzoate oc
5	45	41.3	767	2	T21969	hypothetical prote
6	44	40.4	250	2	G83400	hypothetical prote
7	44	40.4	284	2	F71015	hypothetical prote
8	43.5	39.9	1501	2	T43623	hypothetical prote
9	43	39.4	220	2	C83392	probable glutathio
10	43	39.4	232	2	G75608	hypothetical prote
11	43	39.4	309	2	T32376	hypothetical prote
12	43	39.4	376	2	A48197	opsin, ocellar
13	43	39.4	376	2	B48197	opsin, lateral eye
14	43	39.4	476	2	G64720	probable amino aci
15	43	39.4	476	2	G90629	probable inner mem
16	43	39.4	476	2	G85480	inner membrane tra
17	43	39.4	719	2	S61046	ARPI protein - yea
18	43	39.4	1296	2	T16859	hypothetical prote
19	42.5	39.0	443	2	T21499	hypothetical prote
20	42	38.5	428	2	F81694	hypothetical prote
21	42	38.5	433	2	H87660	peptidoglycan-bind
22	41.5	38.1	345	2	D84012	pyruvate dehydroge
23	41	37.6	128	2	S76955	N-acetylglutamate
24	41	37.6	226	2	A87664	hypothetical prote
25	41	37.6	284	2	A75117	hypothetical prote
26	41	37.6	338	2	T20100	hypothetical prote
27	41	37.6	343	2	B64395	malic acid transpo
28	41	37.6	343	2	T46534	probable FMN-depen
29	41	37.6	347	2	H64371	malic acid transpo

30	41	37.6	419	2	AH3166	hypothetical prote
31	41	37.6	476	2	AG0502	probable amino aci
32	41	37.6	545	2	A87448	conserved hypothet
33	41	37.6	986	1	OTURGA	spectact receptor p
34	41	37.6	2155	2	T30197	alpha tectorin - m
35	40.5	37.2	779	2	T49717	related to PCS1 pr
36	40.5	37.2	1540	2	T45619	hypothetical prote
37	40	36.7	184	2	T35841	probable membrane
38	40	36.7	324	2	T35901	probable arac fami
39	40	36.7	329	2	H70744	hypothetical prote
40	40	36.7	341	2	T35426	probable oxidoredu
41	40	36.7	359	2	T40084	PMWP domain protei
42	40	36.7	390	2	G82844	cysteine synthase
43	40	36.7	430	1	B69009	conserved hypothet
44	40	36.7	462	2	T00708	violaxanthin de-ep
45	40	36.7	468	2	C83160	nitrite extrusion

ALIGNMENTS

RESULT 1

JC7607 prolactin-releasing peptide - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001

C:Accession: JC7607

R:Yamada, M.; Ozawa, A.; Ishii, S.; Shibusawa, N.; Hashida, T.; Ishizuka, T.; Hosoya, Blocham, Biophys. Res. Commun. 281, 53-56, 2001

A:Title: Isolation and characterization of the rat prolactin-releasing peptide gene:

A:Reference number: JC7607; MUID:21092785; PMID:11178959

A:Contents: Spleen

A:Accession: JC7607

A:Molecule type: DNA

A:Residues: 1-83 <YAM>

A:Cross-references: DDBJ:AB040612; DDBJ:AB040613

C:Comment: This peptide induces arachidonic acid metabolite release from rat anterior release, and stimulation of ACTH secretion from the pituitary.

C:Genetics:

A:Gene: PRP

A:Introns: 33/1

Query Match

Best Local Similarity

Matches 18; Conservative

Score 104; DB 2; Length 83;

Pred. No. 5.1e-10;

Mismatches 2; Indels 0; Gaps 0;

DB 33 TPDINPAWYXXRGIRPVGRF 20

33 TPDINPAWYXXRGIRPVGRF 52

RESULT 2

F83376

Conserved hypothetical protein PA2151 [Imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

hypothetical prote
probable amino aci
conserved hypothet
spectact receptor p
alpha tectorin - m
related to PCS1 pr
hypothetical prote
probable membrane
probable arac fami
hypothetical prote
probable oxidoredu
PMWP domain protei
cysteine synthase
conserved hypothet
violaxanthin de-ep
nitrite extrusion

Query Match 45.4%; Score 49.5; DB 2; Length 664;

Best Local Similarity 62.5%; Pred. No. 3.8;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

OY 1 TPDINPAMYXXRGIRP 16
||||| : |
Db 478 TPDINP-WFLORSGRP 492

RESULT 3

T47959

hypothetical protein F15G16.60 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47959

R:De Haan, M.; Maaree, A.C.; Grievell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quet
submitted to the Protein Sequence Database, January 2000

A:Accession: T47959

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1790 <DB>

A:Cross-references: EMBL:AL132959

A:Experimental source: cultivar Columbia; BAC clone F15G16

C:Genetics:

A:Map position: 3

A:Introns: 39/1; 678/2; 698/3; 773/2

A:Note: F15G16.60

Query Match 45.0%; Score 49; DB 2; Length 790;
Best Local Similarity 47.4%; Pred. No. 5.5;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMYXXRGIRVGRF 20
| | | | | : |
Db 366 PDINPRTYGSRLQPHGRW 384

RESULT 4

H82852

hydroxybenzoate octaprenyltransferase XF0068 [imported] - Xylella fastidiosa (strain 9a5

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000

C:Accession: H82852

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82852

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1333 <STB>

A:Cross-references: GB:AE003860; GB:AE003849; NID:99104830; PIDN:AAF82881.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinisch, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Caniarro, L.E.A.; Carraro, D.M.; Carreir, H

as-Neto, E.; Docena, C.; El-Dorry, H.; Facinani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fromm

J.D.; Jungueira, M.L.; Kemp, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurama, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsunuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiter, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0068

C:Superfamily: 4-hydroxybenzoate octaprenyltransferase

Query Match 42.2%; Score 46; DB 2; Length 333;

Best Local Similarity 53.3%; Pred. No. 7.3;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 4 INPAMYXXRGIRPVG 18
::: | | | | |
Db 54 IDPWKRLARGDRPVG 68

RESULT 5

T21969

hypothetical protein F38E11.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21969

R:Matthews, P.
submitted to the EMBL Data Library, January 1996

A:Reference number: Z19495

A:Accession: T21969

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1767 <WIL>

A:Cross-references: EMBL:Z68342; PIDN:CAA92775.1; GSPDB:GN00022; CESP:F38E11.7

A:Experimental source: clone F38E11

C:Genetics:

A:Map position: 4

A:Introns: 50/2; 118/1; 139/2; 189/3; 226/1; 248/1; 287/2; 375/2; 432/3; 465/3; 548/1

Query Match 41.3%; Score 45; DB 2; Length 767;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 PDINPAMY 9
| | | | |
Db 747 PDVKNAMY 754

RESULT 6

G83400

hypothetical protein PA1952 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G83400

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardly, K.; L

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

Nature 406, 959-964, 2000

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: G83400

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1250 <STO>

A:Cross-references: GB:AE004622; GB:AE004091; NID:99947948; PIDN:AG05340.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1952

Query Match 40.4%; Score 44; DB 2; Length 250;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMYXXRGIRPVG 18
| : | | | | : | | | |
Db 233 PALNCAMQDLRALRPSG 249

RESULT 7

F71015

hypothetical protein PH1420 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: F71015

R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: F71015
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-284 <KAW>
A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30526.1; PID:g3257843
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1420
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1420

Query Match 40.4%; Score 44; DB 2; Length 284;
Best Local Similarity 44.4%; Pred. No. 13;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
QY 2 PDINPAMYXXRGIRPVCGR 19
DB 217 PYIEPFYALRGLELLGR 234

RESULT 8
T45623
hypothetical protein F13G24.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45623
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voelt, M.; Robben, J.; Volckaert, G.; Be
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23009
A:Accession: T45623
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1501 <BEV>
A:Cross-references: EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Intons: 64/1; 739/3; 785/2; 1302/2; 1318/3; 1399/2; 1434/2
A:Note: F13G24.180

Query Match 39.9%; Score 43.5; DB 2; Length 1501;
Best Local Similarity 34.5%; Pred. No. 84;
Matches 10; Conservative 5; Mismatches 5; Indels 9; Gaps 2;
QY 1 TPDINPAMYXXRG-----IRPV---GRF 20
DB 1184 SPQMAPSWISQYCTFRKNGLVQPVNDGRF 1212

RESULT 9
C83292
Probable glutathione S-transferase PA2821 [imported] - Pseudomonas aeruginosa (strain PA
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 23-Mar-2001
C:Accession: C83292
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10964043
A:Accession: C83292
A:Molecule type: DNA
A>Status: preliminary
A:Residues: 1-220 <STO>
A:Cross-references: GB:AE004709; GB:AE004091; NID:g9948904; PIDN:AAG06209.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:

A:Gene: PA2821
C:Superfamily: Placoe glutathione transferase

Query Match 39.4%; Score 43; DB 2; Length 220;
Best Local Similarity 64.3%; Pred. No. 15;
Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;
QY 6 PAMYXXRGIRPVCGR 19
DB 39 PAMY--REISPLGR 50

RESULT 10
G75608
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: G75608
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.
.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: AV5250; MUID:20036896; PMID:10567266
A:Accession: G75608
A:Molecule type: DNA
A>Status: preliminary
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12317.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0132
A:Map position: 2

Query Match 39.4%; Score 43; DB 2; Length 222;
Best Local Similarity 46.7%; Pred. No. 16;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 2 PDINPAMYXXRGIRP 16
DB 130 PDHRAAMHLLRGVLP 144

RESULT 11
T32376
hypothetical protein K10F12.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32376
R:Wohlmann, P.; Beck, C.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid K10F12.
A:Reference number: Z21157
A:Accession: T32376
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-309 <WOH>
A:Cross-references: EMBL:AF025462; PIDN:AB71002.1; GSPDB:GN00021; CESP:K10F12.4
A:Experimental source: strain Bristol N2; clone K10F12
C:Genetics:
A:Gene: CESP:K10F12.4
A:Map position: 3
A:Intons: 31/3; 123/2; 196/3; 239/1

Query Match 39.4%; Score 43; DB 2; Length 309;
Best Local Similarity 44.4%; Pred. No. 21;
Matches 8; Conservative 3; Mismatches 5; Indels 2; Gaps 1;
QY 2 PDINPAMYXXRGIRPVCGR 19
DB 131 PDRSPNWTLPKRS--PIGR 146

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 1, 2003, 14:33:06 ; Search time 13.75 seconds

(without alignments)
138.059 Million cell updates/sec

Title: US-09-446-543a-73

Perfect score: 109

Sequence: 1 TPDPINPAMYXKRGIRPVGRFX 22

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	96.3	20	10	US-09-932-161-18
2	105	96.3	31	10	US-09-932-161-15
3	105	96.3	87	12	US-10-044-592-92
4	104	95.4	20	10	US-09-932-161-16
5	104	95.4	20	10	US-09-932-161-17
6	104	95.4	20	12	US-10-044-592-6
7	104	95.4	20	12	US-10-044-592-42
8	104	95.4	21	12	US-10-044-592-43
9	104	95.4	22	12	US-10-044-592-44
10	104	95.4	31	10	US-09-932-161-13
11	104	95.4	31	10	US-09-932-161-14
12	104	95.4	31	12	US-10-044-592-4
13	104	95.4	31	12	US-10-044-592-5
14	104	95.4	31	12	US-10-044-592-39
15	104	95.4	32	12	US-10-044-592-40
16	104	95.4	33	12	US-10-044-592-41
17	104	95.4	70	12	US-10-044-592-80
18	104	95.4	82	12	US-10-044-592-1
19	104	95.4	86	12	US-10-044-592-96

20	104	95.4	91	12	US-10-044-592-94	Sequence 94, Appl
21	104	95.4	98	12	US-10-044-592-28	Sequence 28, Appl
22	104	95.4	98	12	US-10-044-592-38	Sequence 38, Appl
23	104	95.4	98	12	US-10-044-592-82	Sequence 82, Appl
24	104	95.4	98	12	US-10-044-592-84	Sequence 84, Appl
25	104	95.4	98	12	US-10-044-592-86	Sequence 86, Appl
26	104	95.4	98	12	US-10-044-592-88	Sequence 88, Appl
27	98	89.9	19	12	US-10-044-592-27	Sequence 27, Appl
28	93	85.3	29	12	US-10-044-592-76	Sequence 26, Appl
29	83	76.1	25	12	US-10-044-592-78	Sequence 78, Appl
30	57	52.3	9	12	US-10-044-592-8	Sequence 8, Appl1
31	48	44.0	10	12	US-10-044-592-9	Sequence 9, Appl1
32	45	41.3	209	9	US-10-108-915-30	Sequence 30, Appl1
33	44	40.4	276	10	US-09-925-302-535	Sequence 535, App
34	43	39.4	476	10	US-09-815-242-10014	Sequence 10014, A
35	42	38.5	269	9	US-09-738-626-4086	Sequence 4086, Ap
36	42	38.5	604	9	US-09-738-626-3886	Sequence 3886, Ap
37	41	37.6	236	9	US-09-934-392-4	Sequence 4, Appl1
38	41	37.6	417	9	US-09-738-626-4642	Sequence 4642, Ap
39	41	37.6	485	10	US-09-815-242-13712	Sequence 13712, A
40	41	37.6	753	10	US-09-888-615-68	Sequence 68, Appl
41	40	36.7	89	10	US-09-764-877-1227	Sequence 1227, Appl
42	40	36.7	468	10	US-09-815-242-5144	Sequence 5144, Ap
43	39.5	36.2	116	10	US-09-939-625-26	Sequence 26, Appl
44	39.5	36.2	415	10	US-09-924-256A-26	Sequence 26, Appl
45	39	35.8	40	12	US-10-044-592-80	Sequence 80, Appl

ALIGNMENTS

RESULT 1
US-09-932-161-18
Sequence 18, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Civeilli, Olivier
TITLE OF INVENTION: Screening and Therapeutic Methods For
TITLE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 20
TYPE: PRT
ORGANISM: Homo Sapien
US-09-932-161-18

Query Match 96.3%; Score 105; DB 10; Length 20;
Best Local Similarity 90.0%; Pred. No. 5.2e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDPINPAMYXKRGIRPVGRF 20
Db 1 TPDPINPAMYXKRGIRPVGRF 20

RESULT 2
US-09-932-161-15
Sequence 15, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Civeilli, Olivier
TITLE OF INVENTION: Screening and Therapeutic Methods For
TITLE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932.161

;; CURRENT FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 09/560,915
;; PRIOR FILING DATE: 2000-04-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 15
;; LENGTH: 31
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-932-161-15

Query Match 96.3%; Score 105; DB 10; Length 31;
Best Local Similarity 90.0%; Pred. No. 8e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 12 TPDINPAMYASRGIRPVGRF 31

RESULT 3

US-10-044-592-92
;; Sequence 92, Application US/10044592
;; Patent No. US20020143152A1
;; GENERAL INFORMATION:
;; APPLICANT: Hinuma, Shuji
;; APPLICANT: Fukusumi, Shoji
;; TITLE OF INVENTION: Polypeptides, their Production and Use
;; FILE REFERENCE: 2463US2P
;; CURRENT APPLICATION NUMBER: US/10/044,592
;; CURRENT FILING DATE: 2002-01-10
;; PRIOR APPLICATION NUMBER: US 09/403639
;; PRIOR FILING DATE: 1999-25-10
;; PRIOR APPLICATION NUMBER: PCT/JP98/01923
;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: JP 9-109974
;; PRIOR FILING DATE: 1997-04-28
;; NUMBER OF SEQ ID NOS: 96
;; SOFTWARE:
;; SEQ ID NO 92
;; LENGTH: 87
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-044-592-92

Query Match 96.3%; Score 105; DB 12; Length 87;
Best Local Similarity 90.0%; Pred. No. 2.2e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 34 TPDINPAMYASRGIRPVGRF 53

RESULT 4

US-09-932-161-16
;; Sequence 16, Application US/09932161
;; Patent No. US20020037533A1
;; GENERAL INFORMATION:
;; APPLICANT: Clivell, Olivier
;; APPLICANT: Lin, Steven
;; TITLE OF INVENTION: Screening and Therapeutic Methods For
;; TITLE OF INVENTION: Promoting Wakefulness and Sleep
;; FILE REFERENCE: P-UC 4679
;; CURRENT APPLICATION NUMBER: US/09/932,161
;; CURRENT FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 09/560,915
;; PRIOR FILING DATE: 2000-04-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 16
;; LENGTH: 20
;; TYPE: PRT

;; ORGANISM: Bos taurus
US-09-932-161-16

Query Match 95.4%; Score 104; DB 10; Length 20;
Best Local Similarity 90.0%; Pred. No. 7.4e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 5

US-09-932-161-17
;; Sequence 17, Application US/09932161
;; Patent No. US20020037533A1
;; GENERAL INFORMATION:
;; APPLICANT: Clivell, Olivier
;; APPLICANT: Lin, Steven
;; TITLE OF INVENTION: Screening and Therapeutic Methods For
;; TITLE OF INVENTION: Promoting Wakefulness and Sleep
;; FILE REFERENCE: P-UC 4679
;; CURRENT APPLICATION NUMBER: US/09/932,161
;; CURRENT FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 09/560,915
;; PRIOR FILING DATE: 2000-04-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 17
;; LENGTH: 20
;; TYPE: PRT
;; ORGANISM: Rattus
US-09-932-161-17

Query Match 95.4%; Score 104; DB 10; Length 20;
Best Local Similarity 90.0%; Pred. No. 7.4e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 1 TPDINPAMWTGIRPVGRF 20

RESULT 6

US-10-044-592-6
;; Sequence 6, Application US/10044592
;; Patent No. US20020143152A1
;; GENERAL INFORMATION:
;; APPLICANT: Hinuma, Shuji
;; APPLICANT: Fukusumi, Shoji
;; TITLE OF INVENTION: Polypeptides, their Production and Use
;; FILE REFERENCE: 2463US2P
;; CURRENT APPLICATION NUMBER: US/10/044,592
;; CURRENT FILING DATE: 2002-01-10
;; PRIOR APPLICATION NUMBER: US 09/403639
;; PRIOR FILING DATE: 1999-25-10
;; PRIOR APPLICATION NUMBER: PCT/JP98/01923
;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: JP 9-109974
;; PRIOR FILING DATE: 1997-04-28
;; NUMBER OF SEQ ID NOS: 96
;; SOFTWARE:
;; SEQ ID NO 6
;; LENGTH: 20
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; NAME/KEY: PEPTIDE
;; LOCATION: (1)..(20)
;; OTHER INFORMATION: antigen
US-10-044-592-6

Query Match 95.4%; Score 104; DB 12; Length 20;

Best Local Similarity 90.0%; Pred. No. 7.4e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 7
US-10-044-592-42
; Sequence 42, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 24630S2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 42
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-42

Query Match 95.4%; Score 104; DB 12; Length 20;
Best Local Similarity 90.0%; Pred. No. 7.4e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 8
US-10-044-592-43
; Sequence 43, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 24630S2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 43
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-43

Query Match 95.4%; Score 104; DB 12; Length 21;
Best Local Similarity 90.0%; Pred. No. 7.8e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||

Db 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 9
US-10-044-592-44
; Sequence 44, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; APPLICANT: Fukusumi, Shoji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 24630S2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 44
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-44

Query Match 95.4%; Score 104; DB 12; Length 22;
Best Local Similarity 90.0%; Pred. No. 8.1e-10;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 10
US-09-932-161-13
; Sequence 13, Application US/09932161
; Patent No. US20020037533A1
; GENERAL INFORMATION:
; APPLICANT: Civeili, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Screening and Therapeutic Methods for
; FILE REFERENCE: P-UC 4679
; CURRENT APPLICATION NUMBER: US/09/932,161
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/560,915
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-932-161-13

Query Match 95.4%; Score 104; DB 10; Length 31;
Best Local Similarity 90.0%; Pred. No. 1.1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 12 TPDINPAMYAGRGIRPVGRF 31

RESULT 11
US-09-932-161-14
; Sequence 14, Application US/09932161
; Patent No. US20020037533A1
; GENERAL INFORMATION:

APPLICANT: Civeilli, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Screening and Therapeutic Methods for
TITLE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 31
TYPE: PRT
ORGANISM: Rattus
US-09-932-161-14

Query Match 95.4%; Score 104; DB 10; Length 31;
Best Local Similarity 90.0%; Pred. No. 1,1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMWTGIRPVGRF 31

RESULT 12
US-10-044-592-4
Sequence 4, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 4
LENGTH: 31
TYPE: PRT
ORGANISM: Murine
US-10-044-592-4

Query Match 95.4%; Score 104; DB 12; Length 31;
Best Local Similarity 90.0%; Pred. No. 1,1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMWTGIRPVGRF 31

RESULT 13
US-10-044-592-5
Sequence 5, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10

PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 5
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(31)
OTHER INFORMATION: antigen
US-10-044-592-5

Query Match 95.4%; Score 104; DB 12; Length 31;
Best Local Similarity 90.0%; Pred. No. 1,1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMWTGIRPVGRF 31

RESULT 14
US-10-044-592-39
Sequence 39, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 39
LENGTH: 31
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-39

Query Match 95.4%; Score 104; DB 12; Length 31;
Best Local Similarity 90.0%; Pred. No. 1,1e-09;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMWTGIRPVGRF 31

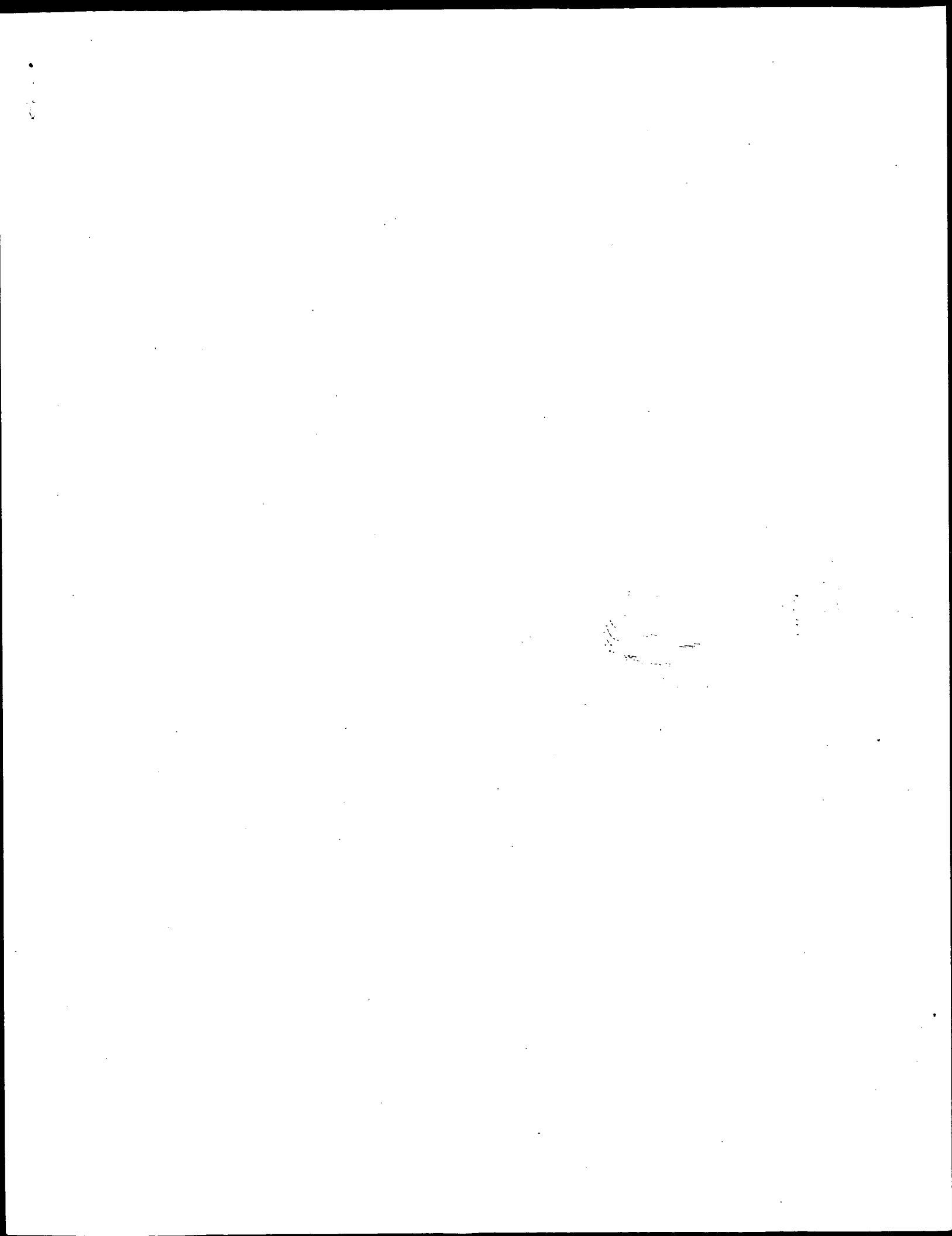
RESULT 15
US-10-044-592-40
Sequence 40, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923

; PRIOR FILING DATE: 1998-04-27
 ; PRIOR APPLICATION NUMBER: JP 9-109974
 ; PRIOR FILING DATE: 1997-04-28
 ; NUMBER OF SEQ ID NOS: 96
 ; SOFTWARE:
 ; SEQ ID NO 40
 ; LENGTH: 32
 ; TYPE: PRT
 ; ORGANISM: Bovine
 US-10-044-592-40

Query Match 95.4%; Score 104; DB 12; Length 32;
 Best Local Similarity 90.08; Pred. No. 1.2e-09;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPYGRF 20
 ||||| |||||
 Db 12 TPDINPAMYAGRGIRPYGRF 31

Search completed: May 1, 2003, 14:46:06
 Job time : 13.75 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:13 ; Search time 10.6562 Seconds
(without alignments)
60.744 Million cell updates/sec

Title: US-09-446-543a-73
Perfect score: 109
Sequence: 1 TPDINPAMYXXRGIRPVGRFX 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	96.3	20	3	US-09-105-678A-46
2	105	96.3	20	4	US-08-776-971-64
3	105	96.3	20	4	US-09-421-208-46
4	105	96.3	20	4	US-09-560-915-18
5	105	96.3	21	3	US-09-105-678A-47
6	105	96.3	21	4	US-08-776-971-65
7	105	96.3	21	4	US-09-421-208-47
8	105	96.3	22	3	US-09-105-678A-48
9	105	96.3	22	4	US-08-776-971-66
10	105	96.3	22	4	US-09-421-208-48
11	105	96.3	31	3	US-09-105-678A-9
12	105	96.3	31	3	US-09-105-678A-43
13	105	96.3	31	4	US-08-776-971-61
14	105	96.3	31	4	US-09-421-208-9
15	105	96.3	31	4	US-09-421-208-43
16	105	96.3	31	4	US-09-560-915-15
17	105	96.3	32	3	US-09-105-678A-44
18	105	96.3	32	4	US-08-776-971-62
19	105	96.3	32	4	US-09-421-208-44
20	105	96.3	33	3	US-09-105-678A-45
21	105	96.3	33	4	US-08-776-971-63
22	105	96.3	33	4	US-09-421-208-45
23	105	96.3	87	4	US-08-776-971-59
24	105	96.3	87	4	US-08-776-971-135
25	105	96.3	87	4	US-08-776-971-138
26	104	95.4	20	3	US-09-105-678A-34
27	104	95.4	20	3	US-09-105-678A-40

28	104	95.4	20	4	US-08-776-971-8	Sequence 8, Appl
29	104	95.4	20	4	US-08-776-971-50	Sequence 50, Appl
30	104	95.4	20	4	US-08-776-971-98	Sequence 98, Appl
31	104	95.4	20	4	US-09-421-208-34	Sequence 34, Appl
32	104	95.4	20	4	US-09-421-208-40	Sequence 40, Appl
33	104	95.4	20	4	US-09-560-915-16	Sequence 16, Appl
34	104	95.4	20	4	US-09-560-915-17	Sequence 17, Appl
35	104	95.4	21	3	US-09-105-678A-35	Sequence 35, Appl
36	104	95.4	21	3	US-09-105-678A-41	Sequence 41, Appl
37	104	95.4	21	4	US-08-776-971-9	Sequence 9, Appl
38	104	95.4	21	4	US-08-776-971-51	Sequence 51, Appl
39	104	95.4	21	4	US-09-421-208-35	Sequence 35, Appl
40	104	95.4	21	4	US-09-421-208-41	Sequence 41, Appl
41	104	95.4	22	3	US-09-105-678A-36	Sequence 36, Appl
42	104	95.4	22	3	US-09-105-678A-42	Sequence 42, Appl
43	104	95.4	22	4	US-08-776-971-10	Sequence 10, Appl
44	104	95.4	22	4	US-08-776-971-52	Sequence 52, Appl
45	104	95.4	22	4	US-09-421-208-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-105-678A-46
; Sequence 46, Application US/09105678A
; Patent No. 6103882
; GENERAL INFORMATION:
; APPLICANT: Suenaga, Masato
; APPLICANT: Moriya, Takeo
; APPLICANT: Tanaka, Yoko
; APPLICANT: Nishimura, Osamu
; TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,678A
; FILING DATE: 26-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 172118/1997
; FILING DATE: 27-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Conlin, David G.
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 46466-342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-105-678A-46
; Query Match 96.3% Score 105; DB 3; Length 20;
; Best Local Similarity 90.0%; Pred. No. 2.9e-11;
; Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TPDINPAMYXXRGIRPVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 2
US-08-776-971-64
Sequence 64, Application US/08776971B
Patent No. 6228984

GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji
Hadata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko

TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
City: Boston

STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 64:

US-08-776-971-64

Query Match 96.3%; Score 105; DB 4; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.9e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXKRGIRPVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 3

US-09-421-208-46
Sequence 46, Application US/09421208
Patent No. 6258561

GENERAL INFORMATION:

APPLICANT: Suenaga, Masato
Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu

TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
City: Boston

STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/421,208

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/105,678

FILING DATE: 26-JUN-1998

APPLICATION NUMBER: JP 172118/1997

FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 48466-342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-421-208-46

Query Match 96.3%; Score 105; DB 4; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.9e-11;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXKRGIRPVGRF 20

Db 1 TPDINPAMYASRGIRPVGRF 20

RESULT 4
US-09-560-915-18

Sequence 18, Application US/09560915
Patent No. 6383764

GENERAL INFORMATION:

APPLICANT: Clevell, Olivier

APPLICANT: Lin, Steven

TITLE OF INVENTION: Therapeutic Compositions and Methods

FILE REFERENCE: P-UC 3534

CURRENT APPLICATION NUMBER: US/09/560,915

CURRENT FILING DATE: 2000-04-28

NUMBER OF SEQ ID NOS: 24

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 18

LENGTH: 20

TYPE: PRT

ORGANISM: Homo Sapien

US-09-560-915-18

Query Match 96.3%; Score 105; DB 4; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.9e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 5

US-09-105-678A-47
Sequence 47, Application US/09105678A
Patent No. 6103882

GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105, 678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-47

Query Match 96.3%; Score 105; DB 3; Length 21;
Best Local Similarity 90.0%; Pred. No. 3.1e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 6

US-08-776-971-65
Sequence 65, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Hadata, Yugo
Kawamata, Yuji

Hosoya, Masaki

Fuji, Ryo

Fukusumi, Shoji

Kitada, Chieko

TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776, 971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 65:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 65:

US-08-776-971-65

Query Match 96.3%; Score 105; DB 4; Length 21;
Best Local Similarity 90.0%; Pred. No. 3.1e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 7

US-09-421-208-47
Sequence 47, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-47

Query Match 96.3%; Score 105; DB 4; Length 21;
Best Local Similarity 90.0%; Pred. No. 3.1e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMVXXRGIRPVGRF 20
DB 1 TPDINPAMVXXRGIRPVGRF 20

RESULT 8
US-09-105-678A-48
Sequence 48; Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-48

Query Match 96.3%; Score 105; DB 3; Length 22;
Best Local Similarity 90.0%; Pred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMVXXRGIRPVGRF 20
DB 1 TPDINPAMVXXRGIRPVGRF 20

RESULT 9
US-08-776-971-66
Sequence 66; Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Habata, Yugo
APPLICANT: Kamata, Yuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 66:

SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-08-776-971-66

Query Match 96.3%; Score 105; DB 4; Length 22;
Best Local Similarity 90.0%; Pred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
||||| |||||
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 10
US-09-421-208-48
Sequence 48, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-48

Query Match 96.3%; Score 105; DB 4; Length 22;
Best Local Similarity 90.0%; Pred. No. 3.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
||||| |||||
DB 1 TPDINPAMYASRGIRPVGRF 20

RESULT 11
US-09-105-678A-9

Sequence 9, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-9

Query Match 96.3%; Score 105; DB 3; Length 31;
Best Local Similarity 90.0%; Pred. No. 4.8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
||||| |||||
DB 12 TPDINPAMYASRGIRPVGRF 31

RESULT 12
US-09-105-678A-43

Sequence 43, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-43

Query Match 96.3%; Score 105; DB 3; Length 31;
Best Local Similarity 90.0%; Pred. No. 4.8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMYASRGIRPVGRF 31

RESULT 13
US-08-776-971-61
Sequence 61, Application US/08776971B
Patent No. 6228984
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
Habata, Yugo
Kawamata, Yuji
Hosoya, Masaki
Fuji, Ryo
Fukusumi, Shoji
Kitada, Chieko
TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,971B
FILING DATE: 06-Feb-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03821
FILING DATE: 28-DEC-1996
APPLICATION NUMBER: JP 7/343371
FILING DATE: 28-DEC-1995
APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996
APPLICATION NUMBER: JP 8/211805
FILING DATE: 12-AUG-1996
APPLICATION NUMBER: JP 8/246573
FILING DATE: 18-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 47176
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 61:
US-08-776-971-61

Query Match 96.3%; Score 105; DB 4; Length 31;
Best Local Similarity 90.0%; Pred. No. 4.8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMYASRGIRPVGRF 31

RESULT 14
US-09-421-208-9
Sequence 9, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-9

Query Match 96.3%; Score 105; DB 4; Length 31;
Best Local Similarity 90.0%; Pred. No. 4.8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMYASRGIRPVGRF 31

RESULT 15
US-09-421-208-43
Sequence 43, Application US/09421208
Patent No. 6258561
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
City: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/421,208
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/105,678
FILING DATE: 26-JUN-1998
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-421-208-43

Query Match 96.3%; Score 105; DB 4; Length 31;
Best Local Similarity 90.0%; Pred. No. 4.8e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
Db 12 TPDINPAMYASRGIRPVGRF 31

Search completed: May 1, 2003, 14:33:54
Job time: 10.6562 secs

1000

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 33.9062 Seconds

(without alignments)
188.386 Million cell updates/sec

Title: US-09-446-543a-61

Perfect score: 171
Sequence: 1 SRTNRHSMETIRPDINPAMYASRGIRPVGRF 31Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	158	92.4	98	6 Q8WN12	Q8WN12 ovis aries
2	98	57.3	117	13 Q9W624	Q9W624 carassius a
3	56.5	33.0	664	16 Q9I1M4	Q9I1M4 pseudomonas
4	55	32.2	315	2 Q87474	Q87474 burkholderi
5	54	31.6	790	10 Q9M371	Q9M371 arabidopsis
6	53.5	31.3	314	16 Q06348	Q06348 mycobacteri
7	53	31.0	327	10 Q94K09	Q94K09 brassica ca
8	53	31.0	327	10 Q94K09	Q94K09 brassica ca
9	53	31.0	327	10 Q94K09	Q94K09 brassica ca
10	52	30.4	785	17 Q9A5E9	Q9A5E9 caulobacter
11	51.5	30.1	420	2 Q93LZ7	Q93LZ7 methanobact
12	51	29.8	54	4 Q9UJF9	Q9UJF9 streptomyce
13	51	29.8	277	11 Q9DBA2	Q9DBA2 mus saplen
14	51	29.8	286	11 Q9DAB5	Q9DAB5 mus musculu
15	51	29.8	450	11 Q9DA19	Q9DA19 mus musculu
16	51	29.8	465	4 Q60687	Q60687 homo sapien

17	51	29.8	465	4 Q6W8B5	Q6W8B5 homo sapien
18	50	29.2	123	2 Q458B3	Q458B3 eubacterium
19	50	29.2	128	16 P74747	P74747 synechocyst
20	50	29.2	692	2 Q9L8U6	Q9L8U6 rhodospirill
21	50	29.2	939	5 Q9GRA0	Q9GRA0 hemicentrot
22	50	29.2	1253	2 Q8RTV7	Q8RTV7 uncultured
23	49.5	28.9	176	3 Q086B9	Q086B9 saccharomyc
24	49	28.7	173	17 Q9EYB3	Q9EYB3 escherichia
25	49	28.7	173	17 Q97A00	Q97A00 thermoplas
26	49	28.7	326	2 Q9F642	Q9F642 stigmatella
27	48.5	28.4	303	16 Q8YWC7	Q8YWC7 anabaena sp
28	48.5	28.4	1501	10 Q9SDB6	Q9SDB6 arabidopsis
29	48	28.1	118	16 Q8UKM3	Q8UKM3 agrobacteri
30	48	28.1	220	16 Q91022	Q91022 pseudomonas
31	48	28.1	335	2 Q88002	Q88002 bordetella
32	48	28.1	335	2 Q45375	Q45375 bordetella
33	48	28.1	348	5 Q45431	Q45431 caenorhabdi
34	48	28.1	455	16 Q33340	Q33340 caenorhabdi
35	48	28.1	514	5 Q961A3	Q961A3 drosophila
36	48	28.1	596	5 Q9VP62	Q9VP62 drosophila
37	48	28.1	863	5 Q9VRV3	Q9VRV3 drosophila
38	48	28.1	1084	5 Q9TYM4	Q9TYM4 caenorhabdi
39	47.5	27.8	184	11 Q8VHS8	Q8VHS8 mus musculu
40	47.5	27.8	273	11 Q8R3B6	Q8R3B6 mus musculu
41	47.5	27.8	333	16 Q9RJ10	Q9RJ10 streptomyce
42	47.5	27.8	401	16 Q8UHH0	Q8UHH0 agrobacteri
43	47.5	27.8	503	16 Q9KRY1	Q9KRY1 vibrio chol
44	47.5	27.8	805	5 Q9TZS3	Q9TZS3 drosophila
45	47	27.5	145	5 Q95ZB5	Q95ZB5 leishmania

ALIGNMENTS

RESULT 1
Q8WN12 PRELIMINARY; PRT; 98 AA.
AC Q8WN12;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Preprolactin-releasing peptide.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Curlew J.D., Kusters D.H.L., Barclay J.L., Anderson S.T.;
RT "Prolactin-releasing peptide (PrP) in the ewe: cDNA cloning, mRNA
RT distribution, and effects on prolactin secretion in vitro and in
RT vivo".
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF450453; AAL47178.1; -; 2A53331ED62CAAB5 CIRC64;
SQ SEQUENCE 98 AA; 10513 MW; 2A53331ED62CAAB5 CIRC64;

Query Match 92.4%; Score 158; DB 6; Length 98;
Best Local Similarity 90.3%; Pred. No. 1, 2e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTNRHSMETIRPDINPAMYASRGIRPVGRF 31
DB 23 SRAHSHMETIRPDINPAMYASRGIRPVGRF 53

RESULT 2

Q8W624 PRELIMINARY; PRT; 117 AA.
AC Q8W624;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE C-RF amide.
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Carassius.
 RN NCBI_TaxID=7957;
 RP SEQUENCE FROM N.A.
 RC TRISSE-BRAIN;
 RA Satate H., Minakata H., Fujimoto M.;
 RT "Carassius kramide (C-RF amide).";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB020024; BAA76662.1;
 SQ SEQUENCE 117 AA; 12879 MW; D5DC4CB2038CZB0 CRC64;

Query Match 57.3%; Score 98; DB 13; Length 117;
 Best Local Similarity 53.8%; Pred. No. 9.4e-07;
 Matches 14; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Db 50 HNDNRSPEDIPFYVGRGVRPIGRF 75

RESULT 3

Q911W4 PRELIMINARY; PRT; 664 AA.
 ID Q911W4
 AC Q911W4;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical protein PA2151.
 GN PA2151.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE-20437337; Pubmed-10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL: AE004642; AG05539.1;
 DR InterPro: IPR001589; Actbind_actin.
 DR InterPro: IPR000461; Alpha_amyase.
 DR Pfam: PF00128; alpha-amyase.1.
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 664 AA; 76329 MW; 8F59FEED54C308AD CRC64;

Query Match 33.0%; Score 56.5; DB 16; Length 664;
 Best Local Similarity 45.8%; Pred. No. 7.6;
 Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

Db 470 YRPFVNTDIPD-WFLQSGRP 492

RESULT 4

Q87474 PRELIMINARY; PRT; 315 AA.
 ID Q87474
 AC Q87474;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Trihydroxytoluene oxygenase.
 GN DNTD.
 OS Burkholderia cepacia (Pseudomonas cepacia).
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
 OC Burkholderia.
 RN NCBI_TaxID=292;
 RP SEQUENCE FROM N.A.
 RC STRAIN-DNT;
 RA Halgler B.E., Johnson G.R., Suen W.C., Spain J.C.;
 RT "Cloning and characterization of Pseudomonas sp. strain DNT genes for
 2,4-dinitrotoluene degradation.";
 RL J. Bacteriol. 175:1831-1837(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DNT;
 RX MEDLINE-93194809; Pubmed-8449889;
 RA Halgler B.E., Johnson G.R., Suen W.C., Spain J.C.;
 RT "Biochemical and genetic evidence for meta-ring cleavage of 2,4,5-
 trihydroxytoluene in Burkholderia sp. strain DNT.";
 RL J. Bacteriol. 181:965-972(1999).
 DR EMBL: AF076848; AAD12738.1;
 SQ SEQUENCE 315 AA; 34745 MW; E99261179022961E CRC64;

Query Match 32.2%; Score 55; DB 2; Length 315;
 Best Local Similarity 44.4%; Pred. No. 5.7;
 Matches 16; Conservative 3; Mismatches 11; Indels 6; Gaps 2;

Qy 1 STHHSHMEIRTPD--INPAMVSRGIRP--VGR 30
 Db 216 SRHHSCFEIDDDAQFLGNKWLASRGWPGWGR 251

RESULT 5

Q9M371 PRELIMINARY; PRT; 790 AA.
 ID Q9M371
 AC Q9M371;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Hypothetical 87.4 kDa protein.
 GN F15616.60.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA De Haan M., Maarse A.C., Grivell L.A., Mewes H.W., Lemcke K.,
 Mayer K.F.X., Quettier F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL132959; CAB71097.1;
 KW Hypothetical protein.
 SQ SEQUENCE 790 AA; 87376 MW; B222724B75650F30 CRC64;

Query Match 31.6%; Score 54; DB 10; Length 790;
 Best Local Similarity 52.6%; Pred. No. 21;
 Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 13 PDINPAMVSRGIRPGRF 31
 Db 366 PPHNPYSGRGLQPHGRW 384

RESULT 6

Q06348

```

ID      006348      PRELIMINARY;      PRT:      314 AA.
AC      O06348;
DT      01-JUL-1997 (TREMBLrel. 04, Created)
DT      01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT      01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE      Hypothetical 33.2 kDa protein (Oxidoreductase, short-chain
DE      dehydrogenase/reductase family).
GN      RV3485C OR MT3589 OR MTCY13E12.38C.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-H37RV:
RC      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R., DeAngelis K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA      Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA      Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA      Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence."
RT      Nature 393:537-544(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-CDC 1551 / OSHKOSH:
RA      Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA      Peterson J., Deboy R., Dodson R., Gwinn M., Hatt D., Hickey E.,
RA      Kolonay J.F., Nelson W.C., Umayam L.A., Ernolova M., Salzberg S.L.,
RA      Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA      Bishai W.;
RT      "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT      laboratory strains."
RT      Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC      -1 (SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC      (SDR) FAMILY.
CC      EMBL; 295390; CAB08708.1; '-'
DR      EMBL; AE007162; AAK47948.1; ALT_INIT.
DR      HSSP; P29132; IDRI.
DR      TIGR; MT3589; '-'
DR      TubercuList; RV3485C; '-'
DR      InterPro; IPR002198; ADH_short.
DR      Pfam; PF00106; adh_short; 1.
DR      PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW      Hypothetical protein; Oxidoreductase; Complete proteome.
SQ      SEQUENCE      314 AA;      33194 MW;      26144BA917E09274 CAC64;

Query Match      31.3%;      Score 53.5;      DB 16;      Length 314;
Best Local Similarity      27.5%;      Pred. No. 9.4;
Matches      11;      Conservative      5;      Mismatches      11;      Indels      13;      Gaps      14.

Db      1      SRTSR-----HSMETRPDINPAMYASRGIRP      27
      1      |||      1      ::      |||
      167      SNTHRMFGAYGTYKSAVDHMKLADDELGSPSVNRNSITRP      226

RESULT 7
ID      094KU9      PRELIMINARY;      PRT:      327 AA.
AC      094KU9;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      Plastid-lipid associated protein PAP1.
OS      Brassica campestris (Field mustard).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Brassica.
OX      NCBI_TaxID=3711;

```

RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=21249173; PubMed=11351096;
RT	"Kim H.U., Wu S.S., Ratnayake C., Huang A.H.; Brassica rapa Has Three Genes That Encode Proteins Associated with Different Neutral Lipids in Plastids of Specific Tissues."
RL	Plant Physiol. 126:330-341(2001).
DR	EMBL: AF290563; AKS7561.1; "
SQ	SEQUENCE 327 AA; 35678 MW; 8C7B87FCD6C02422 CRC64;
OY	4 HRHSMETRPDIPAW 19 :::
Dc	45 HRHDFKVRASDVNDW 60
RESULT 8	
ID	094FZ9 PRELIMINARY; PRT; 327 AA.
AC	094FZ9;
DT	01-DEC-2001 (TREMBLrel. 19, Created)
DI	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE	Plastid-lipid associated protein PAP1.
DS	Brassica campestris (Field mustard).
OC	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eucotylids II; Brassicales; Brassicaceae; Brassica.
OX	NBI_TaxId=3711;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=21249173; PubMed=11351096;
RT	"Kim H.U., Wu S.S., Ratnayake C., Huang A.H.; Brassica rapa Has Three Genes That Encode Proteins Associated with Different Neutral Lipids in Plastids of Specific Tissues."
RL	Plant Physiol. 126:330-341(2001).
DR	EML: AF290566; AKS7564.1; "
SQ	SEQUENCE 327 AA; 35644 MW; 6116E7F1B6C02C88 CRC64;
OY	4 HRHSMETRPDIPAW 19 :::
Dc	45 HRHDFKVRASDVNDW 60
RESULT 9	
ID	09A5E9 PRELIMINARY; PRT; 637 AA.
AC	09A5E9;
DT	01-JUN-2001 (TREMBLrel. 17, Created)
DI	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE	Sensory box histidine kinase/response regulator.
GN	CRC201.
OS	Caulobacter crescentus.
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
CC	Caulobacter.
OX	NBI_TaxId=155892;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 19089 / CB15;
RX	MEDLINE=21173698; Pubmed=11259647;
RA	Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Porocoka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., Darcy R.T., Dodson R.J., Durkin A.S., Gilm M.L., Halt D.H., Kolony J.F., Smith L., Craven M.B., Knowl H., Shetty J., Berry K.,


```

RN [1]
RP SEQUENCE FROM N.A.
RA Lawlor S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035608; CAB55682.1;
FT NON_TER
SQ SEQUENCE 54 AA; 6110 MW; E2F3C39F7B961A9F CRC64;

Query Match
Best Local Similarity 29.8%; Score 51; DB 4; Length 54;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 12 TPDPINPAWYASRGIRP 27
DB 18 TPATPTWYAGSGTYP 33

RESULT 13
O9DBA2 PRELIMINARY; PRT; 277 AA.
AC O9DBA2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 1700023B02Rik protein.
GN 1700023B02Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK008242; BAB2553.1;
DR MGD; MGI:191485; 1700023B02Rik.
DR PRINTS; PR01574; TUBBYPROTEIN.
SQ SEQUENCE 277 AA; 31797 MW; 4B4AD95516FBAC32 CRC64;

Query Match
Best Local Similarity 29.8%; Score 51; DB 11; Length 277;
Matches 14; Conservative 2; Mismatches 8; Indels 6; Gaps 1;

OY 1 SRHRSHMEIRPDINPAWYASRGIRP 30
DB 200 SRHRSHPEKSGDRN-----RGIRSR 223

RESULT 14
O9DAE5 PRELIMINARY; PRT; 286 AA.
AC O9DAE5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)

```

```

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 21, Last annotation update)
DE 1700023B02Rik protein.
GN 1700023B02Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK005900; BAB24307.1;
DR MGD; MGI:191485; 1700023B02Rik.
DR PRINTS; PR01574; TUBBYPROTEIN.
SQ SEQUENCE 286 AA; 32818 MW; 068003CE894827B CRC64;

Query Match
Best Local Similarity 29.8%; Score 51; DB 11; Length 286;
Matches 14; Conservative 2; Mismatches 8; Indels 6; Gaps 1;

OY 1 SRHRSHMEIRPDINPAWYASRGIRP 30
DB 209 SRHRSHPEKSGDRN-----RGIRSR 232

RESULT 15
O9DA19 PRELIMINARY; PRT; 450 AA.
AC O9DA19;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 1700023B02Rik protein.
GN 1700023B02Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK005900; BAB24307.1;
DR MGD; MGI:191485; 1700023B02Rik.
DR PRINTS; PR01574; TUBBYPROTEIN.
SQ SEQUENCE 286 AA; 32818 MW; 068003CE894827B CRC64;

```

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Sessa T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK006260; BAB24488.1; -;
DR MGI; MGI:1914185; 1700023B02Rik.
DR InterPro; IPR000345; CYTC_heme_bind.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
SQ SEQUENCE 450 AA; 5185 MW; F32F11BE6D6A4EAC CRC64;

Query Match 29.8%; Score 51; DB 11; Length 450;
Best Local Similarity 46.7%; Pred. No. 32;
Matches 14; Conservative 2; Mismatches 8; Indels 6; Gaps 1;

QY 1 SRTHRSMEIRTPDINPAMYASRGIRPVGR 30
Db 373 SRSHRSPEKKSQDRN-----RGIRSR 396

Search completed: May 1, 2003, 14:32:53
Job time : 38.9062 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:29:12 ; Search time 17.4375 seconds
(without alignments)
170.906 Million cell updates/sec

Title: US-09-446-543a-61

Perfect score: 171
Sequence: 1 SRTHRHSMETRTPDINPAMYASRGIRPYGRF 31

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : PIR73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	149	87.1	83 2 JC7607	prolactin-releasin
2	56.5	33.0	664 2 F83376	conserved hypothet
3	54	31.6	790 2 T47959	hypothetical prote
4	53.5	31.3	314 2 B70569	hypothetical prote
5	53	31.0	637 2 D87559	sensory box histid
6	52	30.4	785 2 F69099	sensory transducti
7	50.5	29.5	1882 2 S73484	hypothetical prote
8	50	29.2	123 2 S77900	hypothetical prote
9	50	29.2	128 2 S76955	hypothetical prote
10	49.5	28.9	176 2 S67150	hypothetical prote
11	49	28.7	72 2 E91002	hypothetical prote
12	48.5	28.4	303 2 AH2016	probable regulator
13	48.5	28.4	1501 2 T45623	hypothetical prote
14	48	28.1	118 2 AC3169	hypothetical prote
15	48	28.1	220 2 C83292	hypothetical prote
16	48	28.1	335 2 S70671	probable glutathio
17	48	28.1	348 2 T21648	hypothetical prote
18	48	28.1	455 2 D70885	hypothetical prote
19	48	28.1	1084 2 T33759	probable aldC prot
20	47.5	27.8	345 2 AB2664	hypothetical prote
21	47.5	27.8	393 2 AB2664	N-acetylglutamate
22	47.5	27.8	401 2 A97446	conserved hypothet
23	47.5	27.8	503 2 AB2193	hypothetical prote
24	47.5	27.8	533 2 A33111	Sun/nucleolar prot
25	47	27.5	159 2 G82669	segmentation prote
26	47	27.5	215 2 B87577	ubiquitome biosynt
27	47	27.5	284 2 F71015	glutathione S-tran
28	47	27.5	333 2 H82852	hypothetical prote
29	47	27.5	501 2 T48336	hydroxybenzoate oc
				hypothetical prote

30	47	27.5	938 2 C84480	hypothetical prote
31	47	27.5	4589 2 T14914	dynein beta heavy
32	46.5	27.2	240 2 D64688	probable 1-acylgly
33	46.5	27.2	779 2 T49737	related to BCS1 pr
34	46.5	27.2	957 2 AB4089	hypothetical prote
35	46.5	27.2	1495 2 T31434	densin-180 - rat
36	46.5	27.2	1607 2 T13250	hypothetical prote
37	46.5	27.2	1693 1 NNMWHE	genome polypotein
38	46.5	27.2	4957 2 T03455	ALR protein - huma
39	46.5	27.2	5262 2 T03454	ALR protein - huma
40	46	26.9	256 2 F70812	probable lipqr prot
41	46	26.9	342 2 B64395	malic acid transpo
42	46	26.9	347 2 H64371	hypothetical prote
43	46	26.9	419 2 AH3166	hypothetical prote
44	46	26.9	688 2 AT2516	hypothetical prote
45	46	26.9	698 2 T39050	hypothetical prote

ALIGNMENTS

RESULT 1

JC7607 prolactin-releasing peptide - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001

C:Accession: JC7607

R:Yamada, M.; Ozawa, A.; Ishii, S.; Shibusawa, N.; Hashida, T.; Ishizuka, T.; Hosoya, Biochem. Biophys. Res. Commun. 281, 53-56, 2001

A:Title: Isolation and characterization of the rat prolactin-releasing peptide gene.

A:Reference number: JC7607; MUID:21092785; PMID:11178959

A:Contents: Spleen

A:Accession: JC7607

A:Molecule type: DNA

A:Residues: 1-83 <YAM>

A:Cross-references: DDBJ:AB040612; DDBJ:AB040613

C:Comment: This peptide induces arachidonic acid metabolite release from rat anterior

release, and stimulation of ACTH secretion from the pituitary.

C:Genetics:

A:Gene: PRRP

A:Introns: 33/1

Query Match 87.1%; Score 149; DB 2; Length 83;
Best Local Similarity 83.9%; Pred. No. 8.5e-15;
Matches 26; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 SRTHRHSMETRTPDINPAMYASRGIRPYGRF 31

DB 22 SRAHQHSMETRTPDINPAMYRGIRPYGRF 52

RESULT 2

F83376

conserved hypothetical protein PA2151 [Imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83376

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Miroguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yun, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: AB2950; MUID:20437337; PMID:10984043

C:Accession: F83376

A:Molecule type: DNA

A:Residues: 1-664 <STO>

A:Cross-references: GB:AE004642; GB:AE004091; NID:99948163; PIDN:AA05539.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2151

Query Match 33.0%; Score 56.5; DB 2; Length 664;

Best Local Similarity 45.8%; Pred. No. 3.4;
Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

OY 4 HRRSMETPDINPAMVARSICRP 27
DB 470 YRPFVETPDINP-WELGRSR 492

RESULT 3

hypothetical protein F15G16.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47959

R:De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quet
submitted to the Protein Sequence Database, January 2000

A:Reference number: Z24480

A:Accession: T47959

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-790 <DEH>
A:Cross-references: EMBL:AL132959

A:Experimental source: cultivar Columbia; BAC clone F15G16
C:Genetics:

A:Map position: 3
A:introns: 39/1; 678/2; 698/3; 773/2
A:Note: F15G16.60

Query Match 31.6%; Score 54; DB 2; Length 790;
Best Local Similarity 52.6%; Pred. No. 9.7;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 13 PDINPAMVARSICRPVGRF 31
DB 366 PRNPRRTYGSRGQPHGRW 384

RESULT 4

hypothetical protein RV3485c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70569

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.;
Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70569

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-314 <COL>
A:Cross-references: GB:295390; GB:AL123456; NID:93261766; PIDN:CAB08708.1; PID:92104408

A:Experimental source: strain H37RV
C:Genetics:

A:Gene: RV3485c

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
F:46-227/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 31.3%; Score 53.5; DB 2; Length 314;
Best Local Similarity 27.5%; Pred. No. 4.2;
Matches 11; Conservative 5; Mismatches 11; Indels 13; Gaps 1;

OY 1 SRTIR-----HSMETPDINPAMVARSICRP 27
DB 187 SNTIRMGAYGVTKSAVDHMKLAADDELGPVAVNSIRP 226

RESULT 5

DB7559
sensory box histidine kinase/response regulator [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: DB7559

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.;
Laud, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hart, D.H.; Ko
n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: DB7559

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-637 <STO>
A:Cross-references: GB:AE005673; NID:913424056; PIDN:AAK24472.1; GSPDB:GN00148

A:Genetics:

A:Gene: CC2501

Query Match 31.0%; Score 53; DB 2; Length 637;
Best Local Similarity 48.0%; Pred. No. 11;
Matches 12; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 4 HRRSMETPDINPAMVARSICRPV 28
DB 22 HRSDSLRSPALNPAINRVIRIRAV 46

RESULT 6
sensory transduction histidine kinase - *Methanobacterium thermoautotrophicum* (strain
C:Species: *Methanobacterium thermoautotrophicum*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: F69099

R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Oliv, D.; Spadafora, R.; Vitale, C.U.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: fu

A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: F69099

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-785 <MTX>

A:Cross-references: GB:AE000805; GB:AE000666; NID:92621213; PIDN:AMB84680.1; PID:9262

A:Experimental source: strain Delta H
C:Genetics:

A:Gene: MTH174

Query Match 30.4%; Score 52; DB 2; Length 785;
Best Local Similarity 46.4%; Pred. No. 19;
Matches 13; Conservative 1; Mismatches 12; Indels 2; Gaps 1;

OY 4 HRRSMETPDINPAMVARSICRPV 29
DB 412 HRAFRIRPDGNRYRVEYDRPIRTDG 439

RESULT 7
hypothetical protein K05.ori1882 - *Mycoplasma pneumoniae* (strain ATCC 29342)
C:Species: *Mycoplasma pneumoniae*
A:Variety: ATCC 29342

C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73484; S62840

R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li, B.C.; Herrmann, R.

A:Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*

A:Reference number: S73484; MUID:97105885; PMID:8948633

A:Accession: S73484

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1882 <HTM>
A:Cross-references: EMBL:AE000017; GB:U00089; NID:91673812; PIDN:AA95806.1; PID:9167

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1996

A:Accession: AH2016
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1303 <NUR>
A:Cross-references: GB:BA000019, PIDN:BAF78052.1, PID:q17135506, GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1686

Query Match 28.4%; Score 48.5; DB 2; Length 303;
Best Local Similarity 48.4%; Pred. No. 22;
Matches 15; Conservative 1; Mismatches 10; Indels 5; Gaps 2;

QY 4 HRHSMETPDINPAMY---ASRGIRPVGR 30
DB 226 HENHSE-RTRAIDCTYTLTGAGAGGNRPVGR 255

RESULT 13

T45623
hypothetical protein F13G24.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45623
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voelt, M.; Robben, J.; Volckaert, G.; Be
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223009
A:Accession: T45623
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1501 <BEV>
A:Cross-references: EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Insertions: 64/1; 739/3; 785/2; 1302/2; 1318/3; 1359/2; 1434/2
A>Note: F13G24.180

Query Match 28.4%; Score 48.5; DB 2; Length 1501;
Best Local Similarity 29.5%; Pred. No. 1.3e+02;
Matches 13; Conservative 7; Mismatches 9; Indels 15; Gaps 3;

QY 3 THHSMETR-----TPDINPAMTASRG-----IRPV---GRF 31
DB 1169 TKNHSASVADHQDISPOMAPSWYSQYGRFKNGLVCPVNDTGRF 1212

RESULT 14

AC3169
hypothetical protein Atu5077 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AC3169
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kulyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AC3169
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-118 <KUR>
A:Cross-references: GB:AE008687; PIDN:AAI45769.1; PID:q1774503; GSPDB:GN00188
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu5077
A:Genome: plasmid

Query Match 28.1%; Score 48; DB 2; Length 118;
Best Local Similarity 34.4%; Pred. No. 9.3;

Matches 11; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

QY 2 RTRHSMETRPDI-----NPAMYASRGIRPVGR 29
DB 19 KJHKNHMAARRPDICARVQPATPHRGLAGIG 50

RESULT 15

C83292
probable glutathione S-transferase PA2821 [imported] - Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 23-Mar-2001
C:Accession: C83292
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L
; Lorry, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C83292
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-220 <STO>
A:Cross-references: GB:AE004709; GB:AE004091; NID:9948904; PIDN:AG06209.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2821
C:Superfamily: plaiace glutathione transferase

Query Match 28.1%; Score 48; DB 2; Length 220;
Best Local Similarity 44.0%; Pred. No. 18;
Matches 11; Conservative 3; Mismatches 9; Indels 2; Gaps 1;

QY 6 HSMETRPDINPAMTASRGIRPVGR 30
DB 28 YQLEAIAPPGOPAWY--REISPLGR 50

Search completed: May 1, 2003, 14:31:36
Job time : 22.4375 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:33:06 ; Search time 19.375 Seconds
(without alignments)
138.059 Million cell updates/sec

Title: US-09-446-543a-61
Perfect score: 171
Sequence: 1 SRTHRSHMEIRTPDINPAWYASRGIRPVGRF 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues
Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCCT_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	100.0	31	10	US-09-932-161-15
2	171	100.0	87	12	US-10-044-592-92
3	158	92.4	31	10	US-09-932-161-13
4	158	92.4	31	12	US-10-044-592-39
5	158	92.4	32	12	US-10-044-592-40
6	158	92.4	33	12	US-10-044-592-41
7	158	92.4	98	12	US-10-044-592-28
8	158	92.4	98	12	US-10-044-592-28
9	158	92.4	98	12	US-10-044-592-82
10	158	92.4	98	12	US-10-044-592-84
11	158	92.4	98	12	US-10-044-592-86
12	158	92.4	98	12	US-10-044-592-88
13	149	87.1	31	10	US-09-932-161-14
14	149	87.1	31	12	US-10-044-592-4
15	149	87.1	70	12	US-10-044-592-5
16	149	87.1	82	12	US-10-044-592-90
17	149	87.1	86	12	US-10-044-592-1
18	149	87.1	86	12	US-10-044-592-96
19	149	87.1	91	12	US-10-044-592-94

20	147	86.0	29	12	US-10-044-592-26	Sequence 26, Appl
21	128	74.9	25	12	US-10-044-592-78	Sequence 78, Appl
22	113	66.1	20	10	US-09-932-161-18	Sequence 18, Appl
23	109	63.7	20	10	US-09-932-161-16	Sequence 16, Appl
24	109	63.7	20	12	US-10-044-592-42	Sequence 42, Appl
25	109	63.7	21	12	US-10-044-592-43	Sequence 43, Appl
26	109	63.7	22	12	US-10-044-592-44	Sequence 44, Appl
27	105	61.4	20	10	US-09-932-161-17	Sequence 17, Appl
28	105	61.4	20	12	US-10-044-592-27	Sequence 27, Appl
29	103	60.2	19	12	US-10-044-592-80	Sequence 80, Appl
30	88	51.5	40	12	US-10-044-592-80	Sequence 80, Appl
31	57	33.3	9	12	US-10-044-592-8	Sequence 8, Appl
32	54	31.6	209	9	US-10-108-915-30	Sequence 30, Appl
33	54	31.6	428	10	US-09-820-155-4	Sequence 4, Appl
34	50	29.2	428	10	US-09-820-155-2	Sequence 2, Appl
35	48	28.1	10	12	US-10-044-592-9	Sequence 9, Appl
36	46.5	27.2	240	9	US-09-895-913A-184	Sequence 184, Ap
37	46	26.9	417	9	US-09-738-626-4642	Sequence 4642, Ap
38	45.5	26.6	89	10	US-09-764-877-1227	Sequence 1227, Ap
39	45.5	26.6	383	9	US-10-224-446-2	Sequence 2, Appl
40	45	26.3	44	9	US-09-925-299-1094	Sequence 1094, Ap
41	45	26.3	44	10	US-09-925-299-1094	Sequence 1094, Ap
42	45	26.3	236	9	US-09-984-245-318	Sequence 318, Ap
43	45	26.3	236	9	US-09-966-262-318	Sequence 318, Ap
44	45	26.3	236	9	US-09-983-966-318	Sequence 318, Ap
45	45	26.3	236	9	US-10-143-090-318	Sequence 318, Ap

ALIGNMENTS

```

RESULT 1
US-09-932-161-15
: Sequence 15, Application US/09932161
: Patent No. US20020037533A1
: GENERAL INFORMATION:
: APPLICANT: Clivell, Olivier
: TITLE OF INVENTION: Screening and Therapeutic Methods For
: TITLE OF INVENTION: Promoting Wakefulness and Sleep
: FILE REFERENCE: P-UC 4679
: CURRENT APPLICATION NUMBER: US/09/932.161
: PRIOR FILING DATE: 2001-08-17
: PRIOR APPLICATION NUMBER: US 09/560,915
: PRIOR FILING DATE: 2000-04-28
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 15
: LENGTH: 31
: TYPE: PRT
: ORGANISM: Homo Sapien
US-09-932-161-15

Query Match          100.0%   Score 171:   DB 10:   Length 31:
Best Local Similarity 100.0%:   Pred. No. 7.7e-18;
Matches 31:   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY      1 SRTHRSHMEIRTPDINPAWYASRGIRPVGRF 31
      |||
Db      1 SRTHRSHMEIRTPDINPAWYASRGIRPVGRF 31

RESULT 2
US-10-044-592-92
: Sequence 92, Application US/10044592
: Patent No. US20020143152A1
: GENERAL INFORMATION:
: APPLICANT: Hinuma, Shuji
: APPLICANT: Fukusumi, Shoji
: TITLE OF INVENTION: Polypeptides, their Production and Use
: FILE REFERENCE: 2463US2P
: CURRENT APPLICATION NUMBER: US/10/044,592
: CURRENT FILING DATE: 2002-01-10

```

PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 92
LENGTH: 87
TYPE: PRT
ORGANISM: Homo sapiens
US-10-044-592-92

Query Match 100.0%; Score 171; DB 12; Length 87;
Best Local Similarity 100.0%; Pred. No. 2.4e-17;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRTHRSMETRTDPINPAMYASRGIRPVGRF 31
DB 23 SRTHRSMETRTDPINPAMYASRGIRPVGRF 53

RESULT 3
US-09-932-161-13
Sequence 13, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Clivell, Olivier
TITLE OF INVENTION: Screening and Therapeutic Methods for
FILE REFERENCE: P-DC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 31
TYPE: PRT
ORGANISM: Bos taurus
US-09-932-161-13

Query Match 92.4%; Score 158; DB 10; Length 31;
Best Local Similarity 90.3%; Pred. No. 5.6e-16;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTHRSMETRTDPINPAMYASRGIRPVGRF 31
DB 1 SRAHSHMETRTDPINPAMYAGRGIRPVGRF 31

RESULT 4
US-10-044-592-39
Sequence 39, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:

SEQ ID NO 39
LENGTH: 31
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-39

Query Match 92.4%; Score 158; DB 12; Length 31;
Best Local Similarity 90.3%; Pred. No. 5.6e-16;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTHRSMETRTDPINPAMYASRGIRPVGRF 31
DB 1 SRAHSHMETRTDPINPAMYAGRGIRPVGRF 31

RESULT 5
US-10-044-592-40
Sequence 40, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 40
LENGTH: 32
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-40

Query Match 92.4%; Score 158; DB 12; Length 32;
Best Local Similarity 90.3%; Pred. No. 5.8e-16;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRTHRSMETRTDPINPAMYASRGIRPVGRF 31
DB 1 SRAHSHMETRTDPINPAMYAGRGIRPVGRF 31

RESULT 6
US-10-044-592-41
Sequence 41, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 41
LENGTH: 33
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-41

Query Match 92.4%; Score 158; DB 12; Length 33;
 Best Local Similarity 90.3%; Pred. No. 5.9e-16;
 Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDINPAMVYASRGIRPVGRF 31
 DB 1 SRAHSHMETRTDINPAMVYASRGIRPVGRF 31

RESULT 7
 US-10-044-592-28
 ; Sequence 28, Application US/10044592
 ; Patent No. US20020143152A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinuma, Shuji
 ; TITLE OF INVENTION: Polypeptides, their Production and Use
 ; FILE REFERENCE: 2463US2P
 ; CURRENT APPLICATION NUMBER: US/10/044,592
 ; CURRENT FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 09/403639
 ; PRIOR FILING DATE: 1999-25-10
 ; PRIOR APPLICATION NUMBER: PCT/JP98/01923
 ; PRIOR FILING DATE: 1998-04-27
 ; PRIOR APPLICATION NUMBER: JP 9-109974
 ; PRIOR FILING DATE: 1997-04-28
 ; SOFTWARE:
 ; NUMBER OF SEQ ID NOS: 96
 ; SEQ ID NO 28
 ; LENGTH: 98
 ; TYPE: PRT
 ; ORGANISM: Murine
 US-10-044-592-28

Query Match 92.4%; Score 158; DB 12; Length 98;
 Best Local Similarity 90.3%; Pred. No. 1.9e-15;
 Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDINPAMVYASRGIRPVGRF 31
 DB 23 SRAHSHMETRTDINPAMVYASRGIRPVGRF 53

RESULT 8
 US-10-044-592-38
 ; Sequence 38, Application US/10044592
 ; Patent No. US20020143152A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinuma, Shuji
 ; TITLE OF INVENTION: Polypeptides, their Production and Use
 ; FILE REFERENCE: 2463US2P
 ; CURRENT APPLICATION NUMBER: US/10/044,592
 ; CURRENT FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 09/403639
 ; PRIOR FILING DATE: 1999-25-10
 ; PRIOR APPLICATION NUMBER: PCT/JP98/01923
 ; PRIOR FILING DATE: 1998-04-27
 ; PRIOR APPLICATION NUMBER: JP 9-109974
 ; PRIOR FILING DATE: 1997-04-28
 ; NUMBER OF SEQ ID NOS: 96
 ; SOFTWARE:
 ; SEQ ID NO 38
 ; LENGTH: 98
 ; TYPE: PRT
 ; ORGANISM: Bovine
 US-10-044-592-38

Query Match 92.4%; Score 158; DB 12; Length 98;
 Best Local Similarity 90.3%; Pred. No. 1.9e-15;
 Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDINPAMVYASRGIRPVGRF 31
 DB 23 SRAHSHMETRTDINPAMVYASRGIRPVGRF 53

RESULT 9
 US-10-044-592-82
 ; Sequence 82, Application US/10044592
 ; Patent No. US20020143152A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinuma, Shuji
 ; TITLE OF INVENTION: Polypeptides, their Production and Use
 ; FILE REFERENCE: 2463US2P
 ; CURRENT APPLICATION NUMBER: US/10/044,592
 ; CURRENT FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 09/403639
 ; PRIOR FILING DATE: 1999-25-10
 ; PRIOR APPLICATION NUMBER: PCT/JP98/01923
 ; PRIOR FILING DATE: 1998-04-27
 ; PRIOR APPLICATION NUMBER: JP 9-109974
 ; PRIOR FILING DATE: 1997-04-28
 ; NUMBER OF SEQ ID NOS: 96
 ; SOFTWARE:
 ; SEQ ID NO 82
 ; LENGTH: 98
 ; TYPE: PRT
 ; ORGANISM: Bovine
 US-10-044-592-82

Query Match 92.4%; Score 158; DB 12; Length 98;
 Best Local Similarity 90.3%; Pred. No. 1.9e-15;
 Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDINPAMVYASRGIRPVGRF 31
 DB 23 SRAHSHMETRTDINPAMVYASRGIRPVGRF 53

RESULT 10
 US-10-044-592-84
 ; Sequence 84, Application US/10044592
 ; Patent No. US20020143152A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinuma, Shuji
 ; TITLE OF INVENTION: Polypeptides, their Production and Use
 ; FILE REFERENCE: 2463US2P
 ; CURRENT APPLICATION NUMBER: US/10/044,592
 ; CURRENT FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 09/403639
 ; PRIOR FILING DATE: 1999-25-10
 ; PRIOR APPLICATION NUMBER: PCT/JP98/01923
 ; PRIOR FILING DATE: 1998-04-27
 ; PRIOR APPLICATION NUMBER: JP 9-109974
 ; PRIOR FILING DATE: 1997-04-28
 ; NUMBER OF SEQ ID NOS: 96
 ; SOFTWARE:
 ; SEQ ID NO 84
 ; LENGTH: 98
 ; TYPE: PRT
 ; ORGANISM: Bovine
 US-10-044-592-84

Query Match 92.4%; Score 158; DB 12; Length 98;
 Best Local Similarity 90.3%; Pred. No. 1.9e-15;
 Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDINPAMVYASRGIRPVGRF 31
 DB 23 SRAHSHMETRTDINPAMVYASRGIRPVGRF 53

RESULT 11
US-10-044-592-86
; Sequence 86, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 86
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-86

Query Match 92.4%; Score 158; DB 12; Length 98;
Best Local Similarity 90.3%; Pred. No. 1,9e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDPINPAMYASRGIRPVGRF 31
Db 23 SRAHSHMETRTDPINPAMYAGRGIRPVGRF 53

RESULT 12
US-10-044-592-88
; Sequence 88, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 88
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Bovine
US-10-044-592-88

Query Match 92.4%; Score 158; DB 12; Length 98;
Best Local Similarity 90.3%; Pred. No. 1,9e-15;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDPINPAMYASRGIRPVGRF 31
Db 23 SRAHSHMETRTDPINPAMYAGRGIRPVGRF 53

RESULT 13
US-09-932-161-14
; Sequence 14, Application US/09932161
; Patent No. US20020037533A1
; GENERAL INFORMATION:

; APPLICANT: Civel, Olivier
; APPLICANT: Lin, Steven
; TITLE OF INVENTION: Screening and Therapeutic Methods For
; FILE REFERENCE: P-OC 4679
; CURRENT APPLICATION NUMBER: US/09/932,161
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/560,915
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Rattus
US-09-932-161-14

Query Match 87.1%; Score 149; DB 10; Length 31;
Best Local Similarity 83.9%; Pred. No. 1,1e-14;
Matches 26; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDPINPAMYASRGIRPVGRF 31
Db 1 SRAHSHMETRTDPINPAMYTGIRPVGRF 31

RESULT 14
US-10-044-592-4
; Sequence 4, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: JP 9-109974
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Murine
US-10-044-592-4

Query Match 87.1%; Score 149; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 1,1e-14;
Matches 26; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 SRTHRSMETRTDPINPAMYASRGIRPVGRF 31
Db 1 SRAHSHMETRTDPINPAMYTGIRPVGRF 31

RESULT 15
US-10-044-592-5
; Sequence 5, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10

Thu May 1 15:54:16 2003

us-09-446-543a-61.rapb

Page 5

PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 5
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)-(31)
OTHER INFORMATION: antigen
US-10-044-592-5

Query Match 87.1%; Score 149; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.1e-14;
Matches 26; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 SRTHRSHMETPDINPAMVYASGIRPVGRF 31
||:||||| ||||||| |||||||
Db 1 SRAHQSHMETPDINPAMVYASGIRPVGRF 31

Search completed: May 1, 2003, 14:46:06
Job time : 19.375 secs

10

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:36:28 ; Search time 84 Seconds
(without alignments)
80.947 Million cell updates/sec

Title: SEQ74-PLUS-73

Perfect score: 156

Sequence: 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRFX 33

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_21:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.vertebrate:*
15: sp.unclassified:*
16: sp.rvivirus:*
17: sp.bacteriophage:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	93.6	98	6	Q8WN12
2	99	63.5	117	13	Q9W624
3	69	44.2	15281	3	Q09164
4	68	43.6	34350	4	Q8W242
5	67	42.9	7463	16	Q9Z4X6
6	66	42.3	870	4	Q00308
7	66	42.3	870	4	Q96C22
8	66	42.3	870	4	Q9DBH0
9	66	42.3	1889	10	Q94H07
10	66	42.3	4578	13	Q42181
11	66	42.3	26926	4	Q10466
12	66	42.3	26926	4	Q8W2B3
13	65	41.7	2308	5	Q9VDJ9
14	64	41.0	1171	3	Q96U77
15	64	41.0	4135	6	Q18977
16	64	41.0	4247	2	Q9L8H4

17	64	41.0	6620	4	Q96AA2
18	63	40.4	3063	2	Q59497
19	63	40.4	3325	5	Q8T510
20	63	40.4	6889	16	Q8XS40
21	62	39.7	1762	2	Q30480
22	62	39.7	2611	2	Q06487
23	62	39.7	3546	2	Q9F830
24	62	39.7	4133	13	Q9DE12
25	62	39.7	4134	13	Q8QG4
26	62	39.7	5825	10	Q8T731
27	62	39.7	7389	11	Q91Z06
28	62	39.7	8563	2	Q54297
29	61	39.1	565	5	Q9VZV7
30	61	39.1	963	16	Q9CXX2
31	61	39.1	967	2	Q96A7
32	61	39.1	1059	2	Q19370
33	61	39.1	1410	2	Q9K129
34	61	39.1	1771	17	Q8TV64
35	61	39.1	3419	5	Q9GP18
36	61	39.1	3638	4	Q15142
37	61	39.1	4345	5	Q9YIA0
38	61	39.1	4488	11	Q9QZM1
39	61	39.1	5138	10	Q9ZW94
40	61	39.1	6831	5	Q23550
41	61	39.1	7160	5	Q23551
42	61	39.1	7257	2	Q9K127
43	61	39.1	7257	2	Q9K127
44	61	39.1	7525	2	Q9K1E0
45	61	39.1	9477	2	Q9L4X3

ALIGNMENTS

RESULT 1

Q8WN12
ID Q8WN12 PRELIMINARY; PRT; 98 AA.
AC Q8WN12;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Preprolactin-releasing peptide.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Curlew J.D., Kusters D.H.L., Barclay J.L., Anderson S.T.;
RT "Prolactin-releasing peptide (PRP) in the ewe: cDNA cloning, mRNA
RT distribution and effects on prolactin secretion in vitro and in
RT vivo."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF450453; AAL47178.1; --
SQ SEQUENCE 98 AA; 10513 MW; 2A53331ED62CAAB5 CRC64;

Query Match 93.6%; Score 146; DB 6; Length 98;
Best Local Similarity 83.9%; Pred. No. 1e-08;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 23 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 53

RESULT 2
Q9W624 PRELIMINARY; PRT; 117 AA.

AC Q9W624;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE C-RF amide.
 OS Carassius auratus (Goldfish). Craniata: Vertebrata: Euteleostomi:
 OC Eukaryota; Metazoa; Chordata; Teleostei; Ostariophysi; Cypriniformes;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Carassius.
 NC NCBI_Taxid=7957;
 RN [1]
 RP SOURCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Satake H., Minakata H., Fujimoto M.;
 RT "Carassius kramide (C-RF amide).";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB020024; BAA76662.1;
 SQ SEQUENCE 117 AA; 12879 MW; D5DC4CB22038C2B0 CRC64;
 QY Query Match 63.5%; Score 99; DB 13; Length 117;
 Best Local Similarity 39.0%; Pred. No. 0.0014; 7; Indels 10; Gaps 1;
 Matches 16; Conservative 8; Mismatches 7;
 DB 35 SNAHGTVEHDLHIVHVNDRSPEDFVYGVGRVPRIGRF 75
 QY 1 SRXH-----XHSMEKRTPDINPAWXXXRGIRPVGRF 31
 ID 009164 PRELIMINARY; PRT; 15281 AA.
 AC 009164;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Cyclosporin synthetase (CYSPN) (EC 6.---).
 GN SIMA.
 OS Tolypocladium inflatum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae;
 OC Tolypocladium.
 NC NCBI_Taxid=29910;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ATCC 34921;
 RX MEDLINE=95094306; PubMed=8001164;
 RA Weber G., Schoenberger K., Schneider-Scherzer E., Lettner E.;
 RT "The peptide synthetase catalyzing cyclosporine production in
 Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading
 frame.";
 RT Curr. Genet. 26:120-125 (1994).
 RL Cur. Genet. 26:120-125 (1994).
 CC -1- FUNCTION: THE CONSTITUENT AMINO ACIDS OF CYCLOSPORINS ARE
 CC ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS FORMED
 CC THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER INTERMEDIATES.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC -1- COFACTOR: CONTAINS 11 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF CYCLOSPORINS.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN AMP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC EMBL: Z28383; CAAB2227.1;
 DR HSSP; P14687; 1AMU.
 DR InterPro: IPR000873; AMP-Bind.
 DR InterPro: IPR000267; Asp/Glutamase.
 DR InterPro: IPR001242; Condensatn.
 DR InterPro: IPR003880; Pantne_attach.
 DR InterPro: IPR000051; SAM_Bind.
 DR Pfam: PF00501; AMP-binding; 11.
 DR Pfam: PF00668; Condensation; 13.
 DR Pfam: PF00550; pp-binding; 11.
 DR PRINTS: PR00154; AMPBINDING.
 DR PRINTS: PR00139; ASGNLASE.
 DR PRINTS: PS50075; ACP_DOMAIN; 11.
 DR PROSITE: PS00455; AMP_BINDING; 10.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_9.
 KW Lipase; Antilotic biosynthesis; Phosphopantetheine; Repeat;
 KW Multifunctional enzyme.
 FT DOMAIN 15179 15219 13 X 3 AA APPROXIMATE REPEATS.

FT REPEAT 3 1086 DOMAIN 1.
 FT REPEAT 1087 2585 DOMAIN 2.
 FT REPEAT 2586 4072 DOMAIN 3.
 FT REPEAT 4073 5564 DOMAIN 4.
 FT REPEAT 5565 7061 DOMAIN 5.
 FT REPEAT 7062 8121 DOMAIN 6.
 FT REPEAT 8122 9616 DOMAIN 7.
 FT REPEAT 9617 11113 DOMAIN 8.
 FT REPEAT 11114 12185 DOMAIN 9.
 FT REPEAT 12186 13681 DOMAIN 10.
 FT REPEAT 13682 14767 DOMAIN 11 (ALA-ACTIVATING).
 FT BINDING 1060 1060 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 2558 2558 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 4045 4045 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 5537 5537 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 7034 7034 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 8094 8094 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 9589 9589 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 11086 11086 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 12158 12158 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 13654 13654 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 14729 14729 PHOSPHOPANTETHEINE (POTENTIAL).
 SQ SEQUENCE 15281 AA; 1689051 MW; E26DA7AA35324C05 CRC64;
 QY Query Match 44.2%; Score 69; DB 3; Length 15281;
 Best Local Similarity 0.2%; Pred. No. 6.6e+03;
 Matches 20; Conservative 3; Mismatches 8; Indels 11650; Gaps 5;
 QY 1 SRXH-----
 DB 1351 SRHGVTSFAVLLAFAAFAHRLTGSNDATIGVPSANRNPDELNVIGFVNTQICIRITI 1410
 QY 5 -----
 DB 1411 DENDNFESLVRYRSTTTAQNODVPEYQVYSSIMPSSRDASNPVQLMFLHGQOD 1470
 QY 5 -----
 DB 1471 LFKIQLEGTREEVIPTEVTRFEDIEFHYGASKSLGDIIFAADLFEAETIRGVSVFOE 1530
 QY 5 -----
 DB 1531 VLRRLQDPOTPLMTPLTDGIPELERMLHMVKTDPKMSVVDVQOVRLSAETA 1590
 QY 5 -----
 DB 1591 VIDSSRMSYAEIDQRSDQVAAMLROQLPARETFAVLAFRSCAVIALFGILKAGHAYL 1650
 QY 5 -----
 DB 1651 PLDVNPAARLRAILAEVKGKLVLLAGEPSPESQSPVSTVRITADATSPAGHASLDG 1710
 QY 5 -----
 DB 1711 KSKPTAGSLAVYFTSGTGKPKGVMIHGRVLRVYKQTNILSLSPQAOTFERMAHMSILA 1770
 QY 5 -----
 DB 1771 FDASIVEFTALLNGSLVCIDRFITLDAQALEFLREHINIALFPALLKQCTLDAA 1830
 QY 5 -----
 DB 1831 TIKSLDLLVYGGDLDTDAALAKALVKSEVTNAGPTENTVMSITLSTADTERVNGVP 1890
 QY 5 -----
 DB 1891 IGRVNSGVYVMDQNLVPLGVKELVVTGDLARGVTNPALDSRDVDTIANGQLLR 1950
 QY 5 -----
 DB 1951 AVRTGDRARPRPDQGVFFGRMDHQVKVGRHIELAEVEHALLSAGVHADAVVNSQOE 2010
 QY 5 -----

```

Db 2011 DNOGVENAFITTOADNETTLOEAOSSNOVQEWESHFETTAVADITADIDONTLGRDFTSWTS 2070
QY 5 ----- 4
Db 2071 MYDGTLDKREMOBMDJTRTFLDGOAGHVLEIGTGTGMVLFNLQOAGLKSTIGLEPS 2130
QY 5 ----- 4
Db 2131 QSAVOFNKAAQTFPGLEGRAQVGTAMDGRSLSPDLIVINSVAQYFSPREYLAEV 2190
QY 5 ----- 4
Db 2191 VEALVRIQVRRIFPGDMRTYATHKDFLVARAVHTNGSKVRSKVOQEVARLEBEEL 2250
QY 5 ----- 4
Db 2251 VDPAFITLSESEIEHEVELPKNNKVNNELSVRYGAVLHRRNNQNSIRKINA 2310
QY 5 ----- 4
Db 2311 ESWIDFASQMDQGLARLKENKDESIIVENIPYSKTIYERHIAKSLADHDGDTHS 2370
QY 5 ----- 4
Db 2371 SIDGVAMISAREKASQCSLDVHDVQLAEDAGFRVSVSWARSONGALDVFFHHFOP 2430
QY 5 ----- 4
Db 2431 TENESRALVDFPTDYKGOQARSILNRPLOVESRRIEAOVREQLOVLLPAYMIPARIVYL 2490
QY 5 ----- 4
Db 2491 QNNPLNTSGKVDRKELTRAKVTAARTPSESELVAPRDSIEALICEKFKDVLGVEVGITDN 2550
QY 5 ----- 4
Db 2551 FFNVGSHSLATKLARLSQNAQAVNDIFDRPIADIAITODTEHNPILPTSTY 2610
QY 5 ----- 4
Db 2611 GPEQSGFAGRLMFLDOLNVGATWYIMPPAVRLGRLVVSALAAULLAEHERETLRTTF 2670
QY 5 ----- 4
Db 2671 IEQEGIGMOYIHPAPKELAVIDVSGEESTIOKILEEQTPFNLASEPGRALMLKTG 2730
QY 5 ----- 4
Db 2731 EDEHILSTVWHHAISDGWSVDIFQOEIFQFYALLGHDPILAQIAPLSTQYRDFATWQO 2790
QY 5 ----- 4
Db 2791 IFQVAEHRQLAWYTKOLADNKPABLITDFKRPMLSGRAGEIPVVDGLIYEKLODFCR 2850
QY 5 ----- 4
Db 2851 IRQVTAFTVLLAFAFRAHRYMTGTEDATIGTPIANRNPBELGLIGFEVNTQCMRTYDV 2910
QY 5 ----- 4
Db 2911 EDSFETLVHQBRTTLAANODVPEQIVSNILPGSSDTSRNPVLQVLMFALHSQONLCK 2970
QY 5 ----- 4
Db 2971 VRLGIEEBELISIAETTRPDIIEPHLYQEAERLNGSIYAADLFVBETIOSVITIFOGIIQ 3030
QY 5 ----- 4
Db 3031 KGLGEPDMFVASMALDGLJESLRSTGLHPQOTDYPCADSVYOIRKOQVAVNPDIYAVRD 3090
QY 5 ----- 4

```

```

Db 3091 ESTRLSYADLDKRSQDVACMLSRGIAPETFVALLAPRSCETIYAILGVLKANLAVPLD 3150
QY 5 ----- 4
Db 3151 VNPVPSRLBAILSEVSGMLVIGAETPIPEGMAEAFIRITEILLADAKTDINGIAASQ 3210
QY 5 ----- 4
Db 3211 PTBASLAVYIFTSGTGRKGVNVEHKGIVRLTKQTNITSKLPESFHMAHISNLAFDASV 3270
QY 5 ----- 4
Db 3271 WEVFTLLNGTLVCIDYFTLLESTALEKVFPORVNVALPPLAKQCLDNPALVTL 3330
QY 5 ----- 4
Db 3331 SVLYTIGDRLASDAKARGLVQOAFNAYGPTENTVNSTIYPIAEDPFINGVPIGHAVS 3390
QY 5 ----- 4
Db 3391 NSGAFYMDQNOQITPPANGELIYTGDLARGYTTSSINTGRFINVDIDGEOVRAIYFGD 3450
QY 5 ----- 9
Db 3451 RVRYPKDLQIEFGRIDHOVKIRGHRIEPAVEVEYALLSHDLVTDAAVVTSHQENQDLEM 3510
QY 10 ----- 9
Db 3511 VGFVARAVADVEDSSNOVQEWOTHPDSIAYADITTDQOSLGRDFNSWTSMTDGLIK 3570
QY 10 ----- 9
Db 3571 KSOQEWLDDTMRSLDSOPGCHVLEVGTTGMVLFNLGRGGLOSTYGLPSPSATAFV 3630
QY 10 ----- 9
Db 3631 NKAASFPLEDRIVEVGTATNDIDRLGDLHAGLVVNSVAQYFPSQDYLAQLVRDLTK 3690
QY 10 ----- 9
Db 3691 VPGEYRIFPGMRSHAINRDELVARAVHALDCKATKAETQREVVYMESEDELLVDPAFF 3750
QY 10 ----- 9
Db 3751 TSLTQVENIKHVELPRMRATNELSSRYAAVLHVNDLAKPAHKYSPGAWDFATKM 3810
QY 10 ----- 9
Db 3811 DRDALLRLRGTKISDHIAIANIPNSKTIYERTICESYVDLGGDAKDSNDRVSWLSAARS 3870
QY 10 ----- 9
Db 3871 NAVKVASLSAIDLVDIAQAGFRVEISCARQWSONGALDAVFHHLGPOSSHVLIDFLT 3930
QY 10 ----- 9
Db 3931 DHQGRPEALTNHPLHRAQSRVROIERLOTLIPAYMIPAQIMVLDKLPANANGVDR 3990
QY 10 ----- 9
Db 3991 KQLQORAOYVKAQOVSAPVAPRTEIERVLCQEPFSDVLGVDIGIMENFPDLGSHLMATK 4050
QY 10 ----- 9
Db 4051 LAARISRLLETHSVKELFDHPRVCDLVLIQOQSAPHDPIVSTKYTGVPQSGFAGRLM 4110
QY 10 ----- 9
Db 4111 FLDOINFGATWYIMPLAVRLGAMNVHALTAALLALERRHBLRTTFEQNGVMQKVPNP 4170
QY 10 ----- 9
Db 4171 VVTEIRIIDSNQGDYLPITLKEQOTAPFHELETPGWRVALLRLGPGDYILSVVMMHII 4230

```

```

QY 10 ----- 9
Db 4231 SDGMSVDVLEFQELGQFYSTAVKGHDP LSQTPPLHYKDPALMOKKPTQSEHEROLQYW 4290
QY 10 ----- 9
Db 4291 VBQJVD SAPAELLTDLP RPSILSGQAGEMSVTIEGALYKNLEEFCHVHVTSFVLLAAL 4350
QY 10 ----- 9
Db 4351 RAAHYRLTGEDATICTPIANRNPPELEQIIGFFVNTQCI RTVNEDETFESLVOQVNST 4410
QY 10 ----- 9
Db 4411 ATAFPAHODVPEEKIVSTLLPGSRDASRNPVQJMFVHSQKNLGEKLENAHSEVVPTE 4470
QY 10 ----- 9
Db 4471 ITTRFDLEFHLFOODDKLEBSILYSTDLFEAVSOSLSVFEILRRGLNGPDPVPISTLP 4530
QY 10 ----- 9
Db 4531 LODGIYDLOQGLLDVQKTEYPRDSSVVDVFEHQSINPDSIALHGSEKLSYQLDRES 4590
QY 10 ----- 9
Db 4591 DRYARMLRHSFSSDTLIANLAPRSCETIIAFILKANLAVLPDYKAPANIDAIVSS 4650
QY 10 ----- 9
Db 4651 LPGNKILLGANVYPRKIQEAAIDFVPIRDTFTTLTDLQDPTTERPSAOSIAYAMFT 4710
QY 10 ----- 9
Db 4711 SSGTGRPKYVWOHRNIVRLYKNSNVYAKOPAAARIASHLSNLFADASSWEIYAPLNGCA 4770
QY 10 ----- 9
Db 4771 IVCADYTTITDPOALQETFOEHEIRGAMLPRLKQCLVQAPDMISRDLTFAAGDRFS 4830
QY 10 ----- 9
Db 4831 VDALOQRLVSGVFNAVGPTEMTILSTIYVAENDSFVNGVPIGSAVNSGAYIMDKNQ 4890
QY 10 ----- 9
Db 4891 QLVAGVWGLVYTGDLARGYMDPKLDADRFTQJLVNSSEQVRAVYTGDRVYRRPKDQ 4950
QY 10 ----- 13
Db 4951 IEFGRMDQOIKIRGRIEPAVEQAFNDGFVEDVAIVIRTEPNOPEKVAEVTAKGDN 5010
QY 14 ----- 13
Db 5011 SAREEATTOIEGWEAFEGGAYANIEIESEALGYDFMGWTSMTDCTEIDKEMREMLN 5070
QY 14 ----- 13
Db 5071 DTMNSILDGRKAGRVLEVGOTGMIMFNILGRSGLERYIGLEPAPSAEFAVNNAAKSPFG 5130
QY 14 ----- 13
Db 5131 LAGRAEVHGTADVGTLOGITSDMAVINSVAOYFPTPEXLAETIKSLVOVPGMKRIYLG 5190
QY 14 ----- 13
Db 5191 DMRSMA NRDPFAAARAAYSLADNASKDRVAKOMMELEKEBELLYDAFTALASQLODR 5250
QY 14 ----- 13
Db 5251 IQHVEILLPKMKATUNELSSYYAAVHLISDEPLPIYKIDPEAMINFGSRLTREALLAQVL 5310

```

```

QY 14 ----- 13
Db 5311 KENENAESVALSNIPYSKTVERHIVRSIDQEDANNAPEESMDGSDWISAVRTAQCCHTL 5370
QY 14 ----- 13
Db 5371 SASDLFDIAEDAGFRVEVSAROSHOGALDAVEHHLKPATEDSRVLKPEPTDQGRPLK 5430
QY 14 ----- 13
Db 5431 SLTNOPPLPAOSRAELLIREGLQTLPPYMIPOQITLIDRMPINANGKYDRRELRARAK 5490
QY 14 ----- 13
Db 5491 ITQSKPVEDIVPPRNSVEATVCCKFTDVLGVEGITDNFENLGHSLMATKLARLRQ 5550
QY 14 ----- 13
Db 5551 LNTRI SVHDVDPVADLAAVIORNSAPHEPIKPADYTGVPQSFAGRLWFLDQLVNG 5610
QY 14 ----- 13
Db 5611 ATWYIMPLGIRLHGSRLVDALATATAISALQORHREPLRTTFFHEDEGVGVVQDHRPKDLRI 5670
QY 14 ----- 13
Db 5671 IDLSTQPKDAVLAVLKHQTLTFLDLPATEGWRVALIRLGEHEHILSIYMHIIISDGMSVE 5730
QY 14 ----- 13
Db 5731 VLFDEMRFPYSALROODPMQOILLPIQYRDPAAWQKTERQVAEHOROLDYWTGHLADS 5790
QY 14 ----- 13
Db 5791 TPALLTDLPRPSILSRANELPLTIEGRLDKLRACRVHQATPEVILLALRAHYRL 5850
QY 14 ----- 13
Db 5851 TGAEDATIGTPIANRNPBELENMIGFFVNTQCMRIAIEENDNFESLVRVSTATSAFAN 5910
QY 14 ----- 13
Db 5911 QDVPEESIVSSILPGSRASRNPVQJLVLAHVSQODLKLTLBGLRDBAVDSASTFV 5970
QY 14 ----- 13
Db 5971 EFHLEHADRLSGSVLYAKELFKLTIESVVSFELETLRRALDQPLPLAVPLTDGVE 6030
QY 14 ----- 13
Db 6031 IASKGLLDVPTDYPRDANIYEVFOQHVRATPDAIYKDATSILTYAQLDOQSDRLAIWL 6090
QY 14 ----- 13
Db 6091 SRHHMPETLVGLAPRSCETIIAMFGIMKANLAVLPDINSAPARLSILSAVGNLTV 6150
QY 14 ----- 13
Db 6151 LIGSVTAPEDENPEVEAVGIOELLAGTGLDXTGSNARPSATSLAVYIFTSGTGKPKG 6210
QY 14 ----- 13
Db 6211 VAVEHRSVTRLAKPSNVISKLPQGARVAHLANIANIADASIMEIATITLLNGATLVCLDHTTV 6270
QY 14 ----- 13
Db 6271 LDCRTLKEVERESITVTVLMPALLKQCVAEIPETLAIHLLTYTGDRVGGHDAMARSL 6330
QY 14 ----- 13
Db 6331 VRIGMSGYGPTENTVISTIEVDADENFVNGVPIGKTVNSGAYVMDRNOOLVPSGVYG 6390
QY 14 ----- 13

```



```
Db 6391 ELVYTGDLARGYTDPSLANKRPIYITVNGESTIRAYTGTDRVRYRPHDQIEFFGRMDQ 6450
QY 14 ----- 13
Db 6451 VKIRGRIEPEGEVALLSHNSYQDAVAVICAPADDSGAEMVAFVARTDEDEQEE 6510
QY 14 ----- 13
Db 6511 AVDOVGMEHTEFTAAVSEVKDIROSEVNDFMGWTSMYDSEIDKIDMHEMLNDIMRM 6570
QY 14 ----- 13
Db 6571 LDAREBGHVEIGTGMVWFNAKCPGLOGIYGFEPBSKSAQFVDAQSPALDGRS 6630
QY 14 ----- 13
Db 6631 IVHGTATDINKAGPIQPRLVINSVAQYFPEPELFRVEALVQIPSEVRIYFGMRN 6690
QY 14 ----- 13
Db 6691 AINRDVASHALHTLGEKANKRLVROMIYELANEELLTPAFETSLRTLGEKIKHYE 6750
QY 14 ----- 13
Db 6751 ILPKMATNELSKRYRAVLYHGRSREQSTIHQVSPNADIDPAADGLDQOTILNLKEH 6810
QY 14 ----- 13
Db 6811 KDAGTVAIGNIPYSKTIYERFVNKSLSEDDMEGQNSLDGSAMVAVBMAGSCPDLAM 6870
QY 14 ----- 13
Db 6871 DVEIAOEAGYQVEVSMARQWONGALDAIFHHFPPKEGARTLIEPPDYEGRNVTLL 6930
QY 14 ----- 13
Db 6931 NRPLNQSRLGTQIREKLTLLPYMIPSRIMVLDQMPVNNNGKIDKELVRAIYAP 6990
QY 14 ----- 13
Db 6991 KPRSAATRAVPRNEIBAILDEFEDEVIGTEVSLDNFPLGSHLMTKLAARVSHLDA 7050
QY 14 ----- 13
Db 7051 HISIKVFDQPVLAADIASIQRESAPHERIPORPYTGPAQBSFAQRLMFLDQNLGATW 7110
QY 14 ----- 13
Db 7111 YLMPLAIRINGQLRVAALSAALFALERRHETLRTFEESDGVQYIVGEARNSDLRVHDV 7170
QY 14 ----- 13
Db 7171 STGDGEYLEVLRREQTVFPLDSSSEPGWVCLVKTGEEDHVLIVMHHIITDGWVSDLR 7230
QY 14 ----- 13
Db 7231 GELGQYSAALRGQDPLLANPLPIQYRPAAMQREAKOVEHQROLGYSQYVDSIPA 7290
QY 14 ----- 13
Db 7291 ELLTDLPRPSILSGRAGSVDTIEGSYGALQSFCTRSTVTFVLLIVFRIAHFRLFV 7350
QY 14 ----- 13
Db 7351 DDAITGPIANRRPELETVGCFTVNTQCMRISIADDNFEGLVQVRNVAATAAYANDV 7410
QY 14 ----- 13
Db 7411 PFERIVSALVPSRNTSRNPVQVLMFAVQSVEDYDQVRLGLESVMMGEASTRDMETH 7470
QY 14 ----- 13

Db 7471 LVPDQKLGVSILYSSDLFEQGTIONFVDIPEQCLRSVLDQPLRPISVLPFSNAISNLS 7530
QY 14 ----- 13
Db 7531 LDLEMPSTDYPRDRIVVDLFEQQAICPDSTAVKSSQLTYAQLDQSDRAVAMLHER 7590
QY 14 ----- 13
Db 7591 HMPAESLVGLSPRSCETIIAYFGIMKANLAYPLIDVYAPDARLAILDITYGERLLLG 7650
QY 14 ----- 13
Db 7651 AGVPOPGIPIRLSTAYIAELSHATVDVTSIPQSAATSLAYVIFTSGTKPKGVME 7710
QY 14 ----- 13
Db 7711 HNGIYRLVDTNVNVPBSGSALPVSHFSNLAAMDATWEITYAVLNGSTVVCIDRDTMD 7770
QY 14 ----- 13
Db 7771 IALNSTFKENVRAAFETPAFLKQCLAEITPELVANLEILHTAGDRDQGANLAGTKAK 7830
QY 14 ----- 13
Db 7831 GGIFNVLGHTEHTAVSTFYPVVEETFEVNGVYVGRGISNSHAYIIRHQKLVPAVGEL 7890
QY 14 ----- 13
Db 7891 ILTGDGVARGYTDLSANKDRFVYIDINKSTWSYTGDKAKYRPRDQGLEFFGRMDQVK 7950
QY 14 ----- 13
Db 7951 INGVRIEPEVELLTLDBKSVLAATVVRPRPGDEMIATITIDAEDDQVTHKAIYKHL 8010
QY 14 ----- 13
Db 8011 QGILPAMIPSHLYLDDQMPYTDNGKVDKDLALRAQTVQRSTAAVRPRDEVEAVLC 8070
QY 14 ----- 13
Db 8071 EESNULLEVVGITDPFDLGGSHLMTKLAARLSROUNTVSKVDVDPITLADADI 8130
QY 14 ----- 13
Db 8131 RRGSHRDPATPYTGVEQSFQAGRLMFLQNLGASWYIMPAFIRMGPLOTKALAV 8190
QY 14 ----- 13
Db 8191 ALNALVHRHEALRTTFEDHDGVQVYIQPKSSQDLRIITLSDAVIDTAYIALAKREQTTA 8250
QY 14 ----- 13
Db 8251 FDLTSEPGWVSLRLGDDYILSLVMHHIISDGWTVYVLRQELGQYSAIRGQELSQ 8310
QY 14 ----- 13
Db 8311 AKSLPIYRPAVWORQENQIKQAKOLKYSQOLADSTPCEFLTDLPRPSILSGEADAV 8370
QY 14 ----- 13
Db 8371 PMVIGTVYQLLDFPCRTHQVTSFVLLAARTAHYRLTGLDATVGTPIANRRPELEG 8430
QY 14 ----- 13
Db 8431 LIGFEVNTQCMMAISETETPESLVQVQVRLTTEAFANQDVPEQVISTLLPGSRDTSRN 8490
QY 14 ----- 13
Db 8491 PLVOVMFALQSOQDLGRIQLEGMTDEALETPLSRLDLEVLHQLQEVGKLSGLYSTDLF 8550
QY 14 ----- 13
Db 8551 EVETIRGIVDFELIRRGLEQPKQRLMAMPITDGIITKRLDQGLLYAKRAYPRESSVID 8610
```

```

QY 14 ----- 13
Db 8611 LFRQVAAAPDAIAVWDSSTLTYYADLDGOSNKLAMLCORNAPEITLVAFAPRSCULTI 8670
QY 14 ----- 13
Db 8671 VAFGLVKANLAVLPIDVNAFARIEAILSAVPGHKLVLOAHGPELGLTMADTELVOID 8730
QY 14 ----- 13
Db 8731 EALASSSSGDHEQIHASGPTANSIAYMFTSGSTGPKGWMIDHSIIRLVKNSDVAVTL 8790
QY 14 ----- 13
Db 8791 PTEVRMANVNLAFDISVQEIYTFALLNGSTLVCLDYLITLDSKILYNVEAQNAMFT 8850
QY 14 ----- 13
Db 8851 PVLLKOCIGMNPALISRLSVLFNVGDRDLDAHDAVAASGLIODAVYNAYGPTENGQSTWY 8910
QY 14 ----- 13
Db 8911 KDVNEPFGVGPDIGNSTNGAYMDGNOQLVSPGVMEIYVTGDGLARGTDSALDED 8970
QY 14 ----- 13
Db 8971 RFVHTIDGENIKAVRTGDRVRRPKDFIEFFGMDQVKIRGHRIEPAVEHALGH 9030
QY 14 ----- 13
Db 9031 DLVHDAVLRKRPANOEPIEITSOEDETIQHESNKOVOGGEHFDVSRYADIKDD 9090
QY 14 ----- 13
Db 9091 TSFPGHFLGWTSMYDQVDIPVNEKEMLEDETTASILDNRPPGHILEIGAGTMIISNLG 9150
QY 14 ----- 13
Db 9151 KYDGLAKYVGLDPAPSAIFVNEAVKSLPGLAKARVLGTALDGLSDEIQLPELVYI 9210
QY 14 ----- 13
Db 9211 NSVAQYPTSEYLKVVKAVEVPSKRVFFGDIRSOALNRDLAARAVALGDNAKSKED 9270
QY 14 ----- 13
Db 9271 IREKIALESEEBELVDPAFVLSNSQLPNIKHVEVLPKIMKATNELSYRAAVLHIS 9330
QY 14 ----- 17
Db 9331 HNEEQULLIDIDPTAWVDEFAATQKDSQGLRNLQOGRDDVMIAVGNIPYSKTIVERHIM 9390
QY 18 ----- 17
Db 9391 NSLDQDVHNSLDGTSWISDARSAAICTSPDAPALQOLAKEGFRVELSWARORSONGAL 9450
QY 18 ----- 17
Db 9451 DAVHRLATDANCERSRVLVHPTDQGRQLRILTNRLQRAQSRIESQVFALQTLAP 9510
QY 18 ----- 17
Db 9511 AYMIPIRIIVLPQMPNANGKVDKQOLARRAOVAKKRAVSARVAPRNDTEIYCEBYAD 9570
QY 18 ----- 17
Db 9571 ILGTEVGITDNEFDMGHSLMATKLAARLSRRLDTRVTVKEVFDKPYLADLAASIEOGST 9630
QY 18 ----- 17
Db 9631 PHLPJASSVYSGPVEQSYAGQRLMFLDQFNLNATWYHMSLAMRLGLPLNDALDVALRAL 9690

```

```

QY 18 ----- 17
Db 9691 EQRHETLTTFEAKDQIGVYVHEAGMKRLKVLPLSDKNEKEHVALENQMPFTLASE 9750
QY 18 ----- 17
Db 9751 PGWKHLARLGPTXYIISLVNHNHSDGWSVDILROELGOFYSALRGDPLSQVKPLPT 9810
QY 18 ----- 17
Db 9811 QYRPFAMQKFAQVAHEHEROLATWENQADSTPGELLTDEPPRQFLSGKAGYIPVTIEG 9870
QY 18 ----- 17
Db 9871 PYEKILFESKEROYTLFESVLLTFAFRAHFLTGADATIGTPIANRNPBEHIIIGFV 9930
QY 18 ----- 17
Db 9931 NTQCRLLDGTSTFESLVQVRSVAIDAVSNODIPERIVSALLPGSRDASRPLQLM 9990
QY 18 ----- 17
Db 9991 FALHSQDPLGNITTEGLEHERLPTSVATRFMEFHLFOEPKLSGSLIFADELFPETIN 10050
QY 18 ----- 17
Db 10051 SVYTFQELRRLDQPOVSISTMPLDGLIDLEKGLIEISSNFPDYSVVDVFRQOV 10110
QY 18 ----- 27
Db 10111 AANPAPAVNDSMTSMTSLDQKSEQIAMAHLAOGLEPESLIGVAMPRESFTIVSLFGI 10170
QY 28 ----- 27
Db 10171 LKAGYAVLPDVNSPARIQPILEVEGKRLVLLSGIDMPOSDRMVETARIQDILNT 10230
QY 28 ----- 27
Db 10231 KYERSDPMSPSATSLAYIFTSGSTGRPKGMIEHRNILRVQSNVTSQLPQDLRMAH 10290
QY 28 ----- 27
Db 10291 ISNLAFDASIMEITFALLNGCALICIDYFTLLDSQALRTTEKARVNAVTLFAPALLBCL 10350
QY 28 ----- 27
Db 10351 NHAPTLFEDLVLYIGDRLDATDAKIQALVKGTVNAYGPTENTVMSTIYRLTDESY 10410
QY 28 ----- 27
Db 10411 ANGVPIGNAVSSGAYIMDQKORLVPBGVGEVYSGDGLARGYTNTLNDRFVDIVIN 10470
QY 28 ----- 27
Db 10471 DOKARAYRTGDRTRRPKDGSTIEFFGRMDQVKIRGHVREPAVEQAMKKAHDAAYV 10530
QY 28 ----- 27
Db 10531 VOAVDQETEMIGFVMSADRSFEGEEITNOVOMEDHEFESTAVAGIEAIDQATLGRDF 10590
QY 28 ----- 27
Db 10591 TSMYNTYNGNLIDKAMEBEMLDITMOSLLDKEDARPCAEIGTGAMVLEPNLPKNDGLESY 10650
QY 28 ----- 27
Db 10651 VGIERSNAALFVDKAADFPLOQKTLQIVGTADIKLVKDFHDPVVVINSVAQYPPSR 10710
QY 28 ----- 27
Db 10711 SYVOIASLELHMTSVKTIFFGDMRSMATNRDFLVSALYTLGDKATKQJROEVARLEE 10770
QY 28 ----- 27

```

```

Db 10771 NEDELLVPAFFTSILTSQMPGKVKHVELLPKRMRTSNELSSRYAAVLHICRDEGRNRY 10830
QY 28 ----- 27
Db 10831 GRVHSVENAMIDPASSGMDRHALVQMLDERDRAKTVAIGNIPIHSMNTINERHTTSLDT 10890
QY 28 ----- 27
Db 10891 EGEIGADSIDGSAMOSATKAMARCPOLSVTELVEIGQAGFVEVSWARORSQHGLD 10950
QY 28 ----- 27
Db 10951 VVEHHLEDGRVILINFPDPERLPSTGLTSRPLQRIQNRREISOIREQLQTLPLPYM 11010
QY 28 ----- 27
Db 11011 VPSRIIVLERMPLNANSKVKRRELARKARLTQTIKPSATRVAPRNDIEAVLCDEFOAVLG 11070
QY 28 ----- 27
Db 11071 VTVCMDNFEELGHSIMATKLAARLSRRLDTRVSKDIFNOPILDLADVOTGSAPHE 11130
QY 28 ----- 27
Db 11131 AIPSTPYSGVPEQSFSGRLMFLDQMLNMSWYHMPILASRLRGPRLIEALQSALATTEAR 11190
QY 28 ----- 27
Db 11191 HESLRTFERODGVPOIVRAARNKQLRIDVSGTEDAYLAALKQODAFDLTAERQWR 11250
QY 28 ----- 27
Db 11251 VALLRLGPDHVSITVMHHTISDGMSVDILROELGOLYSNASSQAPLPIQYRDFAIWOK 11310
QY 28 ----- 27
Db 11311 QDSQIAEHQOLNWKROLVNSKPABELLADFTPKALSGDADYIPIEIDDQVYQNLRSFC 11370
QY 28 ----- 27
Db 11371 RAHVTSFVALLAFRAAHYRLTGABDATIGSPIANRNREBELGLIGCFVNTQCLRIPVK 11430
QY 28 ----- 27
Db 11431 SEDTFDTLVQARETAEADONDVPEPERIVSSMASSRDTSRNPLVQWFAVHSQHDIG 11490
QY 28 ----- 27
Db 11491 NIRLEGVEKPVSMASSTREDAEMHLFEDQMLGNNVFSKDLFESETIRSVAVAFQETL 11550
QY 28 ----- 27
Db 11551 RRLANPHANLATPLPLTDLPLSLRSLCLOVNOPDPRDASVIDFREQVASIPKSIAYID 11610
QY 28 ----- 27
Db 11611 ASSQLYTELDERSQOLATWLRRQYVPEBEVGYLAPRSCETIARLGIKANLAVLPID 11670
QY 28 ----- 27
Db 11671 VNAPAGRIETILSLPGNRLILIGSDTOAVKLHANSVRFRTISDALVESGSPTEELSTR 11730
QY 28 ----- 27
Db 11731 PTAQSLAYVMEITSGTGVPKGVMEHRTITRLVKNSNVAKOPAAAAIAHLNIAFDASS 11790
QY 28 ----- 27
Db 11791 WEIYAPLNGGVICIDYTTTIDIKALEAVFKOHTRGAMLPALLKQCLVSAPTWISSL 11850
QY 28 ----- 27

```

```

Db 11851 EILFAAGDRLSODALILARRAVSGYYNAYGPTENTVLTINIGENFAFSGVPIGNAY 11910
QY 28 ----- 27
Db 11911 SNSGAFVDMONQOLVSAGVIGELVWTGDGLANGYTDKLRVDRFYITLIDGNRVARYFTG 11970
QY 28 ----- 27
Db 11971 DRYRHRPKDQIEFFFRMDQOIKIRGRIEPAVEQALARDPAISDAVITQLTDEEPE 12030
QY 28 ----- 27
Db 12031 LVAFESLAKNANGTNGVNGVSDQEKIDGDEQHALMKNIRHNLQALLPTYMIPSRITIV 12090
QY 28 ----- 27
Db 12091 DQLPVANGKIDRNEIYVQAQATPRTSSVTVAPRNDIETIICEFADILSVRGITDN 12150
QY 28 ----- 27
Db 12151 FEDIGSHLITATKLAARLSRRLDTRVSRDVFDTVPVQGLASIQOGSTPHEALPALSHS 12210
QY 28 ----- 27
Db 12211 GPVQOSFAQGRMFLDRFNINAAWYIMPEGVRLRGLRYDALQTLRALERHLLRTTF 12270
QY 28 ----- 27
Db 12271 EEOGVGMQIVHSPKRDICVYDIDGANEIDLAKLEEQOAPENLSTEVARVALFKAGEN 12330
QY 28 ----- 27
Db 12331 HHILSIYMHHTISDGMSVDIFQOELAQFYSAVRGHDLISQYKPLPIHYRDPAWQRODK 12390
QY 28 ----- 27
Db 12391 QVAHESOLQYWIEQLADSTPAEILSDFNREVLSEAGTVPYIEDEVYKLSIFCRNH 12450
QY 28 ----- 27
Db 12451 QVTSFVLLAAFRVAHYRLTGABDATIGSPIANRNREBELDIGFVNTQCMRIALEEHD 12510
QY 28 ----- 27
Db 12511 NFLSVRRVSTASAFENODVPEFERIVSALLPGSRDASRNPVQLMFVHSQNLGKIQ 12570
QY 28 ----- 27
Db 12571 LEGLGEPTTYATTTRDVEFHLEFQDKGLAGNVFAADLEEAATIRSVVEVFHEILRG 12630
QY 28 ----- 27
Db 12631 IDQPIAISTMPLVYDGLAALNSRNLPAVEDIEPDFAEASVVDYFQIVVANPDALAVTD 12690
QY 28 ----- 27
Db 12691 TSTKLTYAELOQSDHVAAMLKOKLPAESIYVVLAPRSETIYACIGILKANLAVLPMD 12750
QY 28 ----- 27
Db 12751 SNVPEARQALISEIPGEKFFVLGAGVPIPNKTDVAMVFSIDIVASKTDKSYSPGTRP 12810
QY 28 ----- 27
Db 12811 SASSLAIVITFSGTGRKGVMEHRTISLVKONASRIPQSLMAHVSNLAPASWET 12870
QY 28 ----- 27
Db 12871 FTTLLNGTLECSYFTVLDSKALSAFSDHRINITILPALKQCLADAPSVLSLESL 12930
QY 28 ----- 27
Db 12931 YIGDRLDGADATYKDLVKKAYNAYGPTIENSVMSTIYIIEHETPANGVPIGTSLGPKS 12990

```

QY 28 -----VGRF 31
 DB 12991 KAYIMDQDQQLVAGVAGELVAGDGLANGYDPSLNTGRF 13031
 RESULT 4
 DB 08WZ42 PRELIMINARY; PRT: 34350 AA.
 AC 08WZ42;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
 DE titin.
 GN TTN.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20309627; PubMed=10850961;
 RA Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereousse F.,
 RA Cantner T., Koemerer B., Wilt C., Beckmann J.S., Gregorio C.C.,
 RA Granzier H., Labelt S.;
 RT "Series of exon-skipping events in the elastic spring region of titin
 RT as the structural basis for myofibrillar elastic diversity.";
 RL Circ. Res. 86:1114-1121(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21573839; PubMed=11717165;
 RA Bang M.L., Cantner T., Fornoff F., Geach A.J., Gotthardt M.,
 RA McNabb M., Wilt C.C., Labelt D., Gregorio C.C., Granzier H.,
 RA Labelt S.;
 RT "The complete gene sequence of titin, expression of an unusual -700
 RT kDa titin isoform and its interaction with obscurin identify a novel
 RT Z-line to I-band linking system.";
 RL Circ. Res. 89:1065-1072(2001).
 DB EMBL: AJ277892; CAD12456.1;
 DR InterPro: IPR000282; Cytochrome_c_2.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000577; FGGT_kin.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003599; IG.
 DR InterPro: IPR003598; IG_C2.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR InterPro: IPR002016; Peroxidase.
 DR InterPro: IPR004168; PPAK_molif.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00041; fn3; 132.
 DR Pfam: PF00047; Ig; 146.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF02818; PPAK; 53.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00060; FN3; 133.
 DR SMART: SM00409; IG; 167.
 DR SMART: SM00408; IGG2; 148.
 DR SMART: SM00406; IGV; 23.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TYRc; 1.
 DR PROSITE: PS00933; FGGY_KINASES_1; UNKNOWN_1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
 SQ SEQUENCE 34350 AA; 3816262 MW; 5B1120058A7CE58A CRC64;

Query Match 43.6%; Score 68; DB 4; Length 34350;
 Best Local Similarity 0.1%; Pred. No. 3.1e+04;

Matches 19; Conservative 4; Mismatches 8; Indels 31769; Gaps 5;
 QY 1 SRX-----4
 DB 2505 NRTASDEGPKYKLVGRVETNCNLSVEKIKIIRLRDLTCETQNVVEYELSHSGIDVL 2564
 QY 5 -----4
 DB 2565 WNFKEIKPSSKYYIEAHGKIYKLTVLNMMKDEGKYTFAGENITSGKLTVAGCAISK 2624
 QY 5 -----4
 DB 2625 PLTDQVAESQEAFFCEVAPNDPSKGEWLKDKHLPLTNIRSESGHKRLIATAATKLD 2684
 QY 5 -----4
 DB 2685 DIGETTYVANSKTSIAKLVKAVIKIKTLKMLTETQDAVFYELTHPNKYQWIKNG 2744
 QY 5 -----4
 DB 2745 VLESNEKYAISVKGITYSLRIKNCAYDESVEGRLGLGASARLHETVKIIRKPDV 2804
 QY 5 -----4
 DB 2805 TALENATVAFEVSVSHDTVPVKWFHKNVEIKPSDKHLYSERKYHKMLQNISSDAGEY 2864
 QY 5 -----4
 DB 2865 TAVVQLDECKAKLFVETLITIKTKMNIEVPETKTSFCEVSHFNVPWMKNGVEIEMS 2924
 QY 5 -----4
 DB 2925 EKFKIVVQSKLHLLIMNTSEDSAEYTFVQNDVSATLTPTIMTSMKDINAERKD 2984
 QY 5 -----11
 DB 2985 TTFEYTVYEGISTYKWLKNGVEIKSTDCQKRTKKLHSLNIRNVHGDAADYTFVAGK 3044
 QY 12 -----11
 DB 3045 ANSTATLYVEANHIEFRKHINDIKVLEKKRAMFCEVSEPDITVQMKDQELQITDRIK 3104
 QY 12 -----11
 DB 3105 IQEKYVHRLIPSTRMSDAGKYTVAGGNSTAKLFYEGNDVIRISIKKEVYIEKQRA 3164
 QY 12 -----11
 DB 3165 VVEFEVNEEDVDAHMYKDGIEINPOVERHKYVVERRIHMFISFTRQSDAGEYTFVAGR 3224
 QY 12 -----11
 DB 3225 NRSVTLVYNADEPPQVLOELQPTVTVOSGKPARFCAVISGRPKISWYKEBOLSTGFK 3284
 QY 12 -----11
 DB 3285 CEFLLHDQETTLILLIAPEDAAVYTCERKNDYGVATTSASLSVEVPEVSPDQEMYPY 3344
 QY 12 -----11
 DB 3345 PALITPLDQTVTSEGPAPFCQVSGTDLKYSWYSKDKKIKPSRFRTQFEDTYQLEIA 3404
 QY 12 -----11
 DB 3405 EAVPEDEGYTFVANSNAVGVSTANLSLEAPESILHERIEOIEEMKEKESSEPLSAEE 3464
 QY 12 -----11
 DB 3465 EGIHSAELOLSKINETLELLSSPYSTYFDSSEKSGTGTFIKESVSNADISMGVATLSV 3524
 QY 12 -----11
 DB 3525 TVIGIPKPIQWFFNGVLLTPSADYKVFVPGDGHSLILFTKLEDEGEYTCASNDYGT 3584

QY	12	-----	11
Db	3585	ICSAYLKINSKGEHDKETESAVAKSLEKIGPCPHFLKELKPIRCAQGLPAIFETV	3644
QY	12	-----	11
Db	3645	VGERAPVYTWKREKCOLCTSVYTTIINPNNGSTFIYNDPQREDSGLYICKAENMLGEST	3704
QY	12	-----	11
Db	3705	CAELLVLEEDTDMTPPCAKSTPEAPEDEPQTPLKGRAVELDSQELATFVKDTILK	3764
QY	12	-----	11
Db	3765	AALTEENQOLSYEHIAKANELSSOLPLGAOELQSIIEODKLTPESTREFLCINGSIHQ	3824
QY	12	-----	11
Db	3825	PLKBPSPMLQIYQOSOKTFSKEGILMBEPPEQAVLSDEKTFPSAMSIEQINSITVER	3884
QY	12	-----	11
Db	3885	LKTLAEBEGNYPOSTIEPPMHSTYLSVAEBEVLSPKKEKTVSDTNRQORVTLQKQEAQAL	3944
QY	12	-----	11
Db	3945	ILSGSLAEGHVESLQSPDMISQVNYEPLVPSHSCTEGGKILIESANPLENAGQDSAVR	4004
QY	12	-----	11
Db	4005	IEBKSLRPLALBEKQVLLKEHSDNVMPPOIIESKKEPAIKKQVEOGRDLISKE	4064
QY	12	-----	11
Db	4065	SILSGIPEORLNLKIQICRALQAANAEOPLGFSEWLRNIEKEVEAVNITQEPRIHC	4124
QY	12	-----	11
Db	4125	MYLTSKASVTEEYTTIIEVDPOMANLKMELDALCIIYEETDILTAGSPRIQOAKT	4184
QY	12	-----	11
Db	4185	SLQEMDSFGSQKVEPITEPEVESKYLLISPEVSYENVOGRKYLDATVTGVSASAV	4244
QY	12	-----	11
Db	4245	SDEKQDSLAPSEKESSESSEGEVAATVKIQEABEGFIKEDGPMIHTPLVDVTSBGD	4304
QY	12	-----	11
Db	4305	IYHLLTITINAKENVYEFENKLYPSDEKFKLQDQNTYTLVIDKVTEDHOGGEYCEALN	4364
QY	12	-----	11
Db	4365	DSGKTATSAKLTVKRAAPYIKRIEPLLEVALGHLAKTCEIQSADPNVRQWPKAGREIY	4424
QY	12	-----	11
Db	4425	ESDKSIRSSKYISLEILRTQVDDCEYTCASNEYSVCTATLTVTEAYPPFLSRP	4484
QY	12	-----	11
Db	4485	KSLTFVGRKAKAFICYTGTPIYETITQKDGALSPSPMKISDAENKILLESNLTIQD	4544
QY	12	-----	11
Db	4545	RGVYSCASNNKFGADICQALIIIDRPHFIKLELPVQSAINKKVHLECOVDEDRKVTVM	4604
QY	12	-----	11
Db	4605	SKDGOKLPEGKYKICEBKATLLEIPLAKLKDSTGYVCTASNEAGSSCSATVTVREPP	4664

QY	12	-----	11
Db	4665	SPVKVDPSYIAMPGBSARLHCKLKSPYIQTWFKNNKLEBSNTVRMTFVNSEAILDI	4724
QY	12	-----	11
Db	4725	TDVKVEDSGSYCEAVNDVSDSCSTEIVIKPEPSEIKTLEPADIVRGTNALLQCEVSGT	4784
QY	12	-----	11
Db	4785	GPEIISWFKDKQIRSSKRYRLEFSQKSLVCLIEIFENSADVEYECVANEVKGCNAT	4844
QY	12	-----	11
Db	4845	HLKEPPTFYKVDLLALGGQVTTLQAAVRGSEPISTVMKGQEVIREDKIKMSFSNG	4904
QY	12	-----	11
Db	4905	VAVLIIPDVOISPGKRYTCLAEENAGSQTSGELIVKEPAKIIERAEILOVYAGDPATLE	4964
QY	12	-----	20
Db	4965	YTVAGTPELKFMYKDGRLVASKYRISFKNNVAQLKFTSAELHDSGQYTEISNEYGS	5024
QY	21	-----	20
Db	5025	SSCETTFVLDRIAPFTKPLRNDVSVNGTCRLDCKIAGSLPMRVSFKDKELAASD	5084
QY	21	-----	20
Db	5085	KRIAEVGTASLEIIRVDMNAGNFTCRATNSVSGSKSGALIVQEPSPFYTKGSKOV	5144
QY	21	-----	20
Db	5145	LPBSAVCLKSTQGSTPLIRFKNGKELVSGSCYITKEALESLELYKTSDSGYTT	5204
QY	21	-----	20
Db	5205	CKVSNVAGVEGSANLFWEPATFVEKLEPSQLKKGATOLACKVTGTPPIKITWFAND	5264
QY	21	-----	20
Db	5265	REIKESKHRMSFVESTAVLRLLDVGIEDSGEYCEAONEAGSDHCSSIVIKESPYFTK	5324
QY	21	-----	20
Db	5325	EFKPIELKEVDMLAEVAGTPPEIITWFKDNTILRSGRKKTITQDHLVSLQILKFAV	5384
QY	21	-----	20
Db	5385	ADAGEYQRTNEVSGSISASAVTLREPPSFIKTIESTSLRGTAFAQATLKSLPITV	5444
QY	21	-----	20
Db	5445	TWLKDSDEITEDNIRMTFENNVAASLYLSGLEVKHDKYVCOAKNDAGIQRCSALLSYKE	5504
QY	21	-----	20
Db	5505	PATITEAVSIDVTQGDPAITLVKFSGTKEITAKWFKGQELTSGSKKISVTDVSTLK	5564
QY	21	-----	20
Db	5565	IISTEKDSGEYTFEONDVGRSCKARINVDLIIPSPFKILKKMDSIKSGFIDECI	5624
QY	21	-----	20
Db	5625	VAGSHIPISQWFKDQOELISABEKYFSPHDNTAFLEISQLEGTSGYTCATNKAGHNQ	5684
QY	21	-----	20
Db	5685	CSGHLTVKEPPEYFEKPOSQDVNPNTRVQLKALVGTAPMTIKMFKDKNELHSGAARSVM	5744
QY	21	-----	20

```
Db 5745 KDDTSTLELEPAKADDSGTICQLSNDVGTATSKATLEFKPEPPQIKKPSVLVLRNG 5804
QY 21 ----- 20
Db 5805 STTEPCQITGTPKIRVSWYLDGNEITAIQKHGISFDGLATFOISGARVENSSTYCEAR 5864
QY 21 ----- 20
Db 5865 NDAGTASCISLEKVKPEPTTIRELKVEVVKYSDVELECEVTGTPPEVTWLNKNEIIS 5924
QY 21 ----- 20
Db 5925 SKKVTLDNRVSVFNLHTTKODPSDTGEYQCIVSNEGSSCSTRVAKPEPFIKIENT 5984
QY 21 ----- 20
Db 5985 TTVLKSATQSTVAGSPPISTWLNKDQJLDEDDNVTISFVDSVATLQIRVDNGHSGR 6044
QY 21 ----- 20
Db 6045 YTCQAKNESGVERCYAFLVQEPQAIYEKASVDTEKDMTLECVVAGTPELKVKMLKD 6104
QY 21 ----- 20
Db 6105 GKQIVPSRYFSMPENNVA5FRIO5VMKODSGYTFKENDFGSSGDAYLRVLQNIIP 6164
QY 21 ----- 20
Db 6165 SFTKRLKMDKVLGSSIHMECKVSGSLPISQWFKDGEISTAKYLVCHERSVSLVN 6224
QY 21 ----- 20
Db 6225 NLEEDTANTCKYSNVAGDAGSGLTVKEPSPFLVKGROQAIPOSTVEFKAILKGP 6284
QY 21 ----- 20
Db 6285 PFKIKWFKDVELVSGPCFIGLEGST5FLNYSVDASKTGQYCHVTNDVSDSCITML 6344
QY 21 ----- 20
Db 6345 LVTEPPKRVKLEASKYVACDSSRLCKTAGSPERIVWFRNHELPSADKYMTIDS 6404
QY 21 ----- 20
Db 6405 VAVIQMNLSTEDSGDFICEAONPAGTSCSTKVIKEPVS5FPPIVETLKNAEVSL 6464
QY 21 ----- 20
Db 6465 CELSGTPPEVWYKDKROLSSKRYKIASKNFTSHIILNVDTSDIGYHCKAONEVS 6524
QY 21 ----- 20
Db 6525 DTCVCTVAKPEPPRVSKLNSLTVVAGPAPLQASIRGAOPIFVOWLKEKEVIRESENI 6584
QY 21 ----- 20
Db 6585 RITFENVATLQFAKAPANAGKYICQIKNDGMEENATILVEPAVIEKAPMTVTV 6644
QY 21 ----- 20
Db 6645 GETCTLECKVAGTPEL5VEMTKDGKLLTSQKHKFSYNNKISSLRILISVERQDAGTYTFQ 6704
QY 21 ----- 20
Db 6705 VONNKGSSCTAIVVDSRAVP5FTRLKNTGVLGASCLLECKVAGSSPISVAMPHEK 6764
QY 21 ----- 20
Db 6765 TKIVSAGAKQTTFSDNCTLQILNLSDDMGNYTCVANA5GDECRAVLTVOEP5PEVK 6824
QY 21 ----- 20
Db 6825 EPEPLEVLPGKNVTFTSVIRGTPPEKVM5FRGARELVGDRCNITFEDVVALELEFNIDI 6884
QY 21 ----- 20
Db 6885 SOSGETCVSNNAQASCTTREL5VKEPAPLKLRLSHSVBPKSIIESTYTGLPI5V 6944
QY 21 ----- 20
Db 6945 TWKKDGNITTSKCNITTEKTCILEILNSTKRDAGQYSCIEENAGRDVCGALVSTLE 7004
QY 21 ----- 20
Db 7005 PPFVTELEPLEAAVDSVSIQOQVAGTPEITVSWYKGDTRLRPTPEYRTFTNNVATLV 7064
QY 21 ----- 20
Db 7065 FNKVINDSGETCAENSIGTASSKYFRIOEROLPP5FARQLKDIEQYGLPVYTLGR 7124
QY 21 ----- 20
Db 7125 LNSAPIQCWYRDGVLLEDDENLQTSFVDNVA TLKIQTDL5HSGQYSCSASNP1G7AS 7184
QY 21 ----- 20
Db 7185 SSARLTAREPKSPPEFDIKPV5IDVIAESADECHVTGAQPMRTWSKDKNEIRPGSNY 7244
QY 21 ----- 20
Db 7245 TITCVNTPHLILKAVGKDSGQYTCQATNDYKDMCSAQLSVKPEPKVKKLEASKYAK 7304
QY 21 ----- 20
Db 7305 QGESIQLECKISGPEIKV5FRNDESELHESWKYNSPINSVALLTINEASADSGDYIC 7364
QY 21 ----- 20
Db 7365 EAHNGVGDASCSTALT5KAPVPVFTOKPP5VCAKKSQDVILOCEISGTPPEVWYKDKRQ 7424
QY 21 ----- 20
Db 7425 VRNSKKFKIT5KHFDTSIILNLEASDVGEYHCKATNEVSDTSCSVKFKPEPPFVKL 7484
QY 21 ----- 20
Db 7485 SDT5TLIGDAVELRAIVEGPOPISVWMLKDRGEVIRESENTRISPIDNATLQ5LSPAS 7544
QY 21 ----- 20
Db 7545 NSGKTYICQIKNDAGMECSAVLT5EPARIIEKPEPMYTTGNPPALRCVYTGPELSAK 7604
QY 21 ----- 20
Db 7605 WFKDREL5ADSKHIT5INKVASLKP5CAEM5DKGLY5FEVKN5GK5NCTV5HVS5DR 7664
QY 21 ----- 20
Db 7665 IVP55TRILKQVNA5LIGASV5LECRVSGAPISV5GWFQD5NEI5VSGPKQ5SE5ENCT 7724
QY 21 ----- 20
Db 7725 LNSL5LEP5DTGTYTCVANA5VAG5DECSAVLT5QEP5PE5QT5PDS5EVL5PG5LFT5VI 7784
QY 21 ----- 20
Db 7785 RGT5PP5VKW5K5REL5V5G5CNISL5ED5VTELE5LE5VQ5PLE5G5D5SCL5VTND5AS5C 7844
QY 21 ----- 20
Db 7845 TTH5VKEPAT5VYKRLAD5SVET5GSP5VLEA5YT5GPP5ISV5KID5EYL5ISQ5ERC5ITM 7904
QY 21 ----- 20
Db 7905 TEK5TIL5E5EST5IED5Y5AQ5SCL5IEN5AGQD5IC5EAL5V5LEPP5YFIEPLE5H5EAV5IG5EPA 7964
```

QY 21 ----- 20
Db 7965 TLQCKVDGTEIRISWYKEHTKLRSAPAYKQFKNVNASLVINKVDSDVGEYSCKADNS 8024
QY 21 ----- 20
Db 8025 VGAVASAVLIERKLLPPFPARKLDVHETLGFPAVECRINGSBPLQVSWYKDGVLK 8084
QY 21 ----- 20
Db 8085 DDANLQTSFVHNVAITLIQITDQSHIGQYCSANPLGTASSAKLLISEHEVPPFEDLK 8144
QY 21 ----- 20
Db 8145 PVSVDLALSGTFCCHVTGTAPIKITWAKDNREIRPGNKMVLVENTATLVLKVGK 8204
QY 21 ----- 20
Db 8205 DAGQYCYASNINGKSCSAHLGVQBPPEFIKKLEPSRIYKODEFTRECKIGGSPEIKY 8264
QY 21 ----- 20
Db 8265 LMYKDEIEQSSKFRMSFYDSVAVLEMNLSVEDSGDYTCENHNAAGSASSSTSIKYE 8324
QY 21 ----- 20
Db 8325 PPIFRKKPHPIETLKGADVHLECELOGTBPPIHVSWMYKDRKELSGKKYKIMSENFLLSIH 8384
QY 21 ----- 20
Db 8385 ILNVDAADIGEYCKATNDVGSPTCVGSIALKAPRPFVKLLSDISTVGEKEVOLOTTIEG 8444
QY 21 ----- 20
Db 8445 AEPISVWERKGEIVRESDNWISYSENIATLQFSRVEPANAKYTCQIKNDAGQWECF 8504
QY 21 ----- 20
Db 8505 ATLSVLEPATIVKPEISIVTTGDTCTLECTVAGPELSTKWKFKDKELTSDNKYKISPF 8564
QY 21 ----- 20
Db 8565 NKVSGKIINVAAPSDGYSFVEQNVPVKDSCASLQVSDRTVPSPFTKKLETNGLSGS 8624
QY 21 ----- 20
Db 8625 SYVMECKYSGSPPISVSMFHGNEISSGRKYQTILLDNTCALTYMMLSESDGDYTCIAT 8684
QY 21 ----- 20
Db 8685 NMAGSDECSAPLTVREPPSFVOKPDPMDVLTGTNVTFTSIKGTBPFSVSWKGSSELP 8744
QY 21 ----- 20
Db 8745 GDRCNVLEDSVALELFDVDTSGSEYTCIIVSNEAGKASTTHLYIKAPAKFVRINDY 8804
QY 21 ----- 20
Db 8805 STIEKGPILLEGFTGTPEISVTWKNGINWTPSORCNITTEKSAILEIPSIVEDAQO 8864
QY 21 ----- 20
Db 8865 YNCIENMNSGKSCSAQILILEPPIFVAKOLEPVKVSGDSASLQCLAGTPEIGVSWYK 8924
QY 21 ----- 20
Db 8925 DTKLRPTTYKMHFRNNVATLVFNQVDINDSGEYICKAENSVEYASASTFLVQOKLPP 8984
QY 21 ----- 20
Db 8985 SFSROLRDVQETVGLPVYFDCAISGSEPISVSWYKDGKPLKDSPNVQTSFLDNTATLNI 9044

QY 21 ----- 20
Db 9045 KIDRSLAGQYSCATNPISGASSARLLITBGNPPFIDIRLAPDAVAGSADFECHVT 9104
QY 21 ----- 20
Db 9105 GTQPIKSWAKDSREIRSGKQYISYLENSAHLTVLKYDKGDSGQYTCYAVNEVGKDSCT 9164
QY 21 ----- 20
Db 9165 AOLNIKERLIPSPFTKRLSEVETEENGSFKLEGRVAGSOPITVAMYKNNIEIQTSNCE 9224
QY 21 ----- 20
Db 9225 IFFKNNTLVQYRKAGMDAGLYTCVSDNAGSALCTSSIVIKKEPKPPVDOHLTPVTY 9284
QY 21 ----- 20
Db 9285 SEGEYVOLSCHVQSEPIRIQWLKAGREIKPSDRCSFASGTAVLELDVAKADSGDY 9344
QY 21 ----- 20
Db 9345 CKASNVASDPTTKSVTTIKDPAPAPATKKAVDGRLEFVSEPOSIRVEXTATATIAKY 9404
QY 21 ----- 20
Db 9405 GGDPIPNVTKWKGRQOLNOGRVFIHOKGDEAKLEIPIOTTKTDSGLKYCAFNHGBIE 9464
QY 21 ----- 20
Db 9465 SNVLQVDERKKOEKIEGDLRALMKTPILKKGAGEEIDIMELKNVDPKEYEYARM 9524
QY 21 ----- 20
Db 9525 YGIDPFRLLQAFELKQSEETHRLIEIEIENSEDEKEFEELVSFTQORLSQTEPVT 9584
QY 21 ----- 20
Db 9585 LIKDIEQTVLKNDAVEIDIKINYPEIKLSWKGTENKLEPSKPELSDGDRHTLRV 9644
QY 21 ----- 20
Db 9645 NCOLKQGNRYLVCGRPHIASAKLTVIEPAMERHLODVLKEGQCTMQCFSPVNVKSEW 9704
QY 21 ----- 20
Db 9705 FRNGRIILKPOGRHKTVEHKKYKLTIDVRAEDOGQTTCKEEDLETSABLRIEAPIQFT 9764
QY 21 ----- 20
Db 9765 KRIONIVSEHQSATFCEVSPDDAIVTWYKGPTELDESQKYNFRNDGRCHYMTIHNVTP 9824
QY 21 ----- 20
Db 9825 DDEGYSVIARLEPRGEARSTAEILYLTKEIKELKLPDIPDSKVPITPPIRAVPEEL 9884
QY 21 ----- 20
Db 9885 PPVAPPILPLPPEEKPPPKRIEYTKKAVKDKDAKKVAKREKMTPREIYKPPPPPT 9944
QY 21 ----- 20
Db 9945 TLIPAKAPEIIDVSSKAEEVKIMTITRKKEVOKKEEAVYERKQAVHKEKRVFTESPEPY 10004
QY 21 ----- 20
Db 10005 DELEVEPYTEPFEQPYEEDEDEYEELKVAKKEVHEMEDEFEEOEYEREDEGE 10064
QY 21 ----- 20
Db 10065 EEMEAYQEREVIOVQKEVEEESHERRKVPKVPKAPPPPVKIKKPVLEKIKTSRRHE 10124
QY 21 ----- 20

Db 10125 EKKVQTKVPEVSKKIVPOKPSRTVQOEVIIEKVPAVHTKKMVISSEKMFASHTSEEV 10184
QY 21 ----- 20
Db 10185 SVTPVEQKEIVTEEKIHAVSKRVEPPKVPDELPEKPAEBAVAPVIPKKEVPPAKVP 10244
QY 21 ----- 20
Db 10245 EYPPKKVPEKKKVPVKKKEPPAKPPKVPPEKKIVPVPVAKKKKEAPPAKVPQK 10304
QY 21 ----- 20
Db 10305 RVVTEEKITIVQRESPPPAVPEIPKKKVPEERKVPKKEEVPVPPKVPALPKKVPDE 10364
QY 21 ----- 20
Db 10365 EKVAVPVPAKKAPPPRAVSKKTVVEEKRFVAEKLSPAVHQVETRHEVSAEHEMSY 10424
QY 21 ----- 20
Db 10425 SEEBEGSISVYREEREEREEBAVTEYEMMEPEEYVEEKLHISKVEAPPAVTER 10484
QY 21 ----- 20
Db 10485 QEKKIVLPKIPAKIEBPPPAKVPEAPKKIVPEKKVPADVPKKEKVPVPPKVEEPKVPV 10544
QY 21 ----- 20
Db 10545 EKKVPKVIKMEBPLPAKVTERRHQITOEKVLVAVTKEAPPAKAVPEEPKRAVEEKV 10604
QY 21 ----- 20
Db 10605 LKLPKREEPKPAKVIFFRRKRVKKEKVSIEAPKREPOPIKEVTIMEEKRAVTLLEBAV 10664
QY 21 ----- 20
Db 10665 SVOREEYEEYEEYDYEKEPEEYEPTEYDQYEEYERREYERHEBYITPEKPIPVKP 10724
QY 21 ----- 20
Db 10725 VPPEPVTCKRAPPAKVLKKAPEEKVPVPIPKKLPKPPKVEEKKVVEEKIRISIK 10784
QY 21 ----- 20
Db 10785 REKQVTEPAKAVPMKRVAAEKEVPVPRKEVAPVAVPEVPEKELEPEEVAESEEVTIH 10844
QY 21 ----- 20
Db 10845 VEELYVEEHEBYIHHEEEFTTEEEVVPVIVKVPVPRKVPPEKKKVPVPKKKEAPPAK 10904
QY 21 ----- 20
Db 10905 VPPEVPPKPEEKVPVLIKKKKPPPAKVPVPPKVPPEEKVPVPPKKEVAPPAKVPVPR 10964
QY 21 ----- 20
Db 10965 KPVEKKVPVPAKVEAPPAKVPPEVKLIPSEKKTVPYKKEVAPPPVPKKEBPV 11024
QY 21 ----- 20
Db 11025 PVALPOEEVILFEELIVEPEEVLPEEEVILPEEEVILPEEEVILPEEEVILPEEEVILPE 11084
QY 21 ----- 20
Db 11085 EBYVPEEEVYPEEVLPEYKPKVPVAPVPEIKKVTKEKKVIVPKKEAPPAKVPVPR 11144
QY 21 ----- 20
Db 11145 KVEEKRIILPEEEVILPEVTEPEEPEEPISEEIPEEPPSIEVEEVAAPPVPEVIKAV 11204
QY 21 ----- 20

Db 11205 PEAPTPVPPKKEVAPPAKVSKKIPEEKVPVPOKKEAPPAVPEVPKVPKKEVLPKKEA 11264
QY 21 ----- 20
Db 11265 VPPAKGRVLEEKVSVAFRQEVVVKERLELVVAEVEELPEEEHEHEVEEYEEGEFHE 11324
QY 21 ----- 20
Db 11325 VEEFKLEQHREBEHREKVRHVIIEVFAEVEVEFEKPRAPKGEPISEKIIPKPKPT 11384
QY 21 ----- 20
Db 11385 KVPKKEBPAKVPVPPKIVVEKRVVDEBRVPTKVPDVLPEKEVVPKVPVPPAK 11444
QY 21 ----- 20
Db 11445 PEAPPPKVPPEAPKEVPEKKVPVPPKPEVPTKVPPEVPAKAVEKKVPEALIPKPESP 11504
QY 21 ----- 20
Db 11505 PPEVPEAKVEVPEKKVPAPBPKKPEVTPVKVPEAPKEVPEKKVPVPPKKEVPPTKV 11564
QY 21 ----- 20
Db 11565 PEVPAVPEKKVPPEALIPKPESPPEVTEPEEVEVALBEPPAEVVEEPBAAPPQVTPP 11624
QY 21 ----- 20
Db 11625 KKPVEKKAAPAVALKPELPLPVKVPVPEVPEKVPVLPVPPKKEAPPAKVPPEVPEV 11684
QY 21 ----- 20
Db 11685 PEKKAVALKEVPPAKVPEVPPKRVLEKPAVPEPESAESPPEVYEEPELAPEEELIA 11744
QY 21 ----- 20
Db 11745 PEEKRVVAEEBEPVPPPAVPEEPKIIPEKKVPVIAKKEAPPPKPEPEKVIKPKL 11804
QY 21 ----- 20
Db 11805 KRPVPPPPAPKEDVKKEKIFOLKAIPPKKVPEKQVPEKVELPLKVGGEKKVRKLLP 11864
QY 21 ----- 20
Db 11865 EKKPEKEEVLKSVLRKPEEPEKVPKLEKVKPAVPEPPPKVEEVEVPVTKR 11924
QY 21 ----- 20
Db 11925 EKKIPEPTKVPKIPAIPLPAPEPKRPEAEVKTIKPPVEPEPTPIAAPTVPVVGKA 11984
QY 21 ----- 20
Db 11985 EAKAPKEAAKPKGPIKGVPKTPSIEAERKRLPGSGEKPDEAFTYOLKAVPLKF 12044
QY 21 ----- 20
Db 12045 VKEIKDIIITSEFVSSAIEECIVSPSTAITTMKDSNIRESPKHFADGDKRLHT 12104
QY 21 ----- 20
Db 12105 IDVQSDAGEYTCVLRIGNKEKSTAKLVEELPVRFVKTLEEVTVVKGQPLVISCEN 12164
QY 21 ----- 20
Db 12165 KERDVWRKDKIYVEKGRIVPGVIGMRAITINDADDTAGITVTVENANNIECSC 12224
QY 21 ----- 20
Db 12225 VAVVEYIRDMLVKPIRDQHVKKGTAFACDIADKTPAIKWFKGDEIPABNDKTILR 12284
QY 21 ----- 20
Db 12285 DGNHLVLIKINAMPEDIAEYAVEIGKRYPAKLTLGERREVELLKPIEDVTIYEKSASF 12344


```

QY 21 ----- 20
Db 12345 AEISEADIPGOMKLGELLRSPCEIKAEGRKRLTLRKVKLDQAGEVLQALNATTA 12404
QY 21 ----- 20
Db 12405 ILTYKEIELDAVPLKDVTPERRQARFECVLTREANYMSKGPDIKSDKFDIADGK 12464
QY 21 ----- 20
Db 12465 KHLIVINDSQFDDSGVYTAAYEKGKTSARLFTVGRKFKMSPLEDQYVKESETATVCEL 12524
QY 21 ----- 20
Db 12525 SHEKMHVWFKNDAKLHTSRTVLISSEKTKHLEMKEVTLDDISQIKAYVELSSTAOLK 12584
QY 21 ----- 20
Db 12585 VLEADPYFTVKLHDKTAVEKDEITLKCEVSKDVPVKWKGEEIYVSPKYSIKADGLRI 12644
QY 21 ----- 20
Db 12645 LKIKKADLKDGEEVCOGDTKTANTVEARLIKVEKPLXGVEVFGETAHEIEISEP 12704
QY 21 ----- 20
Db 12705 DVHGOMLKGPPLTASPDCEIIEDGKHHILHNCOLGTMGEVSFOANAKSAANLKVKE 12764
QY 21 ----- 20
Db 12765 LPLIFIRPLSOVKYFEKDEAKFECEVSRERPTFRWLKGTQETIGDDRFELIKDGTKHSW 12824
QY 21 ----- 20
Db 12825 IKSAAFEDEAKYFEADBKHTSGKLIIEGIRLKFPLTKDVTAKESAVETVELSHDN 12884
QY 21 ----- 20
Db 12885 RYKFKNDQRLHTTRSVSMODEGKTHSTFRKDISIDTSOIRVEAMGMSSEAKLTVELEGD 12944
QY 21 ----- 20
Db 12945 PYFTGLQDYTGVEKDEVILQCEISKADAPVKWFKDGKEIKPSNNAYIKADGKKRMILK 13004
QY 21 ----- 20
Db 13005 KALKSDIGQYTCDCGTGKTSKGLDIEDREIKLVPLHVSVEMETEFARFETISEDDIHA 13064
QY 21 ----- 20
Db 13065 NWKLKGBALLQTPDCEIKESGKTHSLVLANCRILDQGTGVDFOANAKSSAHLRVPKRVIG 13124
QY 21 ----- 20
Db 13125 LLEPLKDVYTAGETATFDELSYEDIPVEMYLKGGKLEPSDKVVRSEKGVHTLIRDV 13184
QY 21 ----- 20
Db 13185 KLEADAGEVOLTARDFKTHANLFVKEPVEPTKPLEDQYVEGATVALECEVSHENAKVKM 13244
QY 21 ----- 20
Db 13245 FKNGTETLKSKEYIADVGRKRLVJHDCIPEDIKITYTCDAKCFKTSCLNINVPPHVEPL 13304
QY 21 ----- 20
Db 13305 RPLTLDQVREKEMARFECESRENAKVKWFKDGAELKKGKVDIISKGAIRLIVINKCL 13364
QY 21 ----- 20
Db 13365 DDEAETSCERVARTSGMLTVLEEAVFTKNLANIEVSETDTIKLCEVSKPGAEVIWTK 13424

```

```

QY 21 ----- 20
Db 13425 GDEEIIETGVEYELITGSKRKLIVIONAHLLEDAGNVCRLPSSRTDGKRVHLEAEFISK 13484
QY 21 ----- 20
Db 13485 PONLEILEGEKAEFVCSISKESFPVQMKRDKTLESGDKYDVYIADGKKRVLYVKDATTOD 13544
QY 21 ----- 20
Db 13545 MGYVYVWVGARAARAAHLTVIEKLRIYVPLKDPYRVKEQOEVENCEVNTGEGAKAKFERNEE 13604
QY 21 ----- 20
Db 13605 AIFDSSKYLIILOKDLVYTLIRDAHLDDOANTVSLTNRGENVASANLIVEEDBLRTV 13664
QY 21 ----- 20
Db 13665 EPLKDIETMEKKSVTWCKVNLNLTAKTKNGEEVFPDNRVSYRYDKYKHLITIKDGF 13724
QY 21 ----- 20
Db 13725 PDEGEYITVAGODKVAEHLIIETAPFEVHELEDQTVTEBDAAVSCOLSREKANVKWYR 13784
QY 21 ----- 20
Db 13785 NGREIKGKKYKFKENDGSIHRLIKDCRLDDECEYACGVEDKSRARLFEIEIPEIIRP 13844
QY 21 ----- 20
Db 13845 PODILEPAGADVFLAELNKDKVEYOMLRNNVVYVGDKHOMSEGIHRLQICDIKPRD 13904
QY 21 ----- 20
Db 13905 QGEYRIANDKEARAKLELAAPKIKTADQDLVDVKGPLTVWVPDYAPKAEAEKREN 13964
QY 21 ----- 20
Db 13965 EPLSTWIDTAEQTSFRILEAKKGDKGRYKIVLONKHGKAGFINLKYIDVPGVNRLE 14024
QY 21 ----- 20
Db 14025 VYETFDGEVSLAMEEPLDNGSGKIIGYVVERDIRKKTWVLTADRAESCEFTVGLQKGG 14084
QY 21 ----- 20
Db 14085 VEYLFYVSARNRVGCEPEVETDNPVEARSKYDVPBPPLANTVITTDVNRGVSILTWEPPEYD 14144
QY 21 ----- 20
Db 14145 GGAETITNVYIELRDKTSIRMDTAMTVRAEDLSAFTVDVVEGOEYSFRVAQNRIGVKPS 14204
QY 21 ----- 20
Db 14205 AATPYKVAADPIERPSPVNLTSDDQTOSSVOLKWEPPKDGSEILGYITIERCEEGDN 14264
QY 21 ----- 20
Db 14265 WJRCNMKLVPELTYKVTLGKGNKLYLVSANENKAGVSDPSIIGPLTADDAFVEPTMDL 14324
QY 21 ----- 20
Db 14325 SAFKDGLEVIYBNPJTIIIVPSTGYPRPATWCFGDKVLETGDRVMKTLISAYAEVLSPS 14384
QY 21 ----- 20
Db 14385 ERSDKGIITLKENVKTIISGEIDVNVJARPSAPKELFGDITDSVHLTWEPDDOGS 14444
QY 21 ----- 20
Db 14445 PLTGYVVEKREYSRKTWTKVMDFVTDLBFTYVDLVQGEYLFKVCARNKCGPGRAYVDE 14504
QY 21 ----- 20

```

Db 14505 PVNMSTPAIVDPDPENVKMDRTANSIFLTMWPKNDGSGRIKGYIVERCPRGSDKMWAC 14564
QY 21 ----- 20
Db 14565 GEPVAETKMEVTELEBGKWYAVRVKALNRGASKPSRPTETEIQAVDTQAPAEFIIDVKLL 14624
QY 21 ----- 20
Db 14625 AGLTVAAGTIELPAVTVGKPEPKITWTAKADMLKODKRITTEINVPKSTVTIIVDSKSD 14684
QY 21 ----- 20
Db 14685 TGYIIEAVVCGRAVAVEVNVLDKPPPAFDITDVNESCLTMNPBDGSGSKITN 14744
QY 21 ----- 20
Db 14745 YVVERRATDSEVWHKLSYVKDTNFKATKLIPIKKEYIFRVAENMYGVGEVQASBITAK 14804
QY 21 ----- 20
Db 14805 YQEDPPEPPTRIEBSDITKDAVTLTWCEPDGSGPTTGWVERLBDPDDKWKRCMKMY 14864
QY 21 ----- 20
Db 14865 KDTTYRVKGLTNKKKYRFLAENLAGPRKSKSTEPILIKDIPDPWPGKPTVKDVGK 14924
QY 21 ----- 20
Db 14925 TSVRLMNTKPEHDGAKIESYVIEMLKTGDEVNVAEGVPTTOHLLPGIMEQEXSEFY 14984
QY 21 ----- 20
Db 14985 RAVKAGESESESESDVLCREKLYPSPRMLEVINITNTADIKMTVEKDGSGPTIN 15044
QY 21 ----- 20
Db 15045 YIVERKDVRRKMQVTDTVTKDCTVTPLEGSLYFRVAENALGQSDYTEIEDSVLA 15104
QY 21 ----- 20
Db 15105 KDTTTPGPPYALAANDVTKRHNDLKWPPKNDGSRPIQRYIEKKERLGRVYKAKGTA 15164
QY 21 ----- 20
Db 15165 GPDCNFRATDVIESTEVQFQVRAENAGVGHSEPTIELSIEDPTSPSPDLHVTDAQ 15224
QY 21 ----- 20
Db 15225 RKHIAIAKPPKNGSGSPIIGYHVMCPVTEKWMRVNSRPIDKPKVEGVVPDKEYV 15284
QY 21 ----- 20
Db 15285 LRVAVNAIGVSESEISENVAVADPOCKPTIDLETHDIIVIEGKLSIVPPRAVVPVT 15344
QY 21 ----- 20
Db 15345 VSMHKGKEVAKASDRLTMKNDHISAHLEVPKSVRADAGIYITILENKLGSATASINVKI 15404
QY 21 ----- 20
Db 15405 GLPGCKDIKASDITKSSCKLTWEPPEPDGTPILHYVLERREAGRTIYIPVMSGENKLS 15464
QY 21 ----- 20
Db 15465 WTVKDLIPNGEYFFRVKAVNVGVGEYIELKNVIAODPKOPRDPVDEVHNPTAEMT 15524
QY 21 ----- 20
Db 15525 ITWKPPLYDGSGKIMGYIIEKIAAGEERMKRCNEHLVPLITYTAKGLEGEKYOFRVRAE 15584
QY 21 ----- 20

Db 15585 NAAGISESRATPPTKAVDPIDAPKVLILRTSLVEKRDEIDALDASISGSPYITTIWIKD 15644
QY 21 ----- 20
Db 15645 NVTPEEIKKRAAPLVRRRKGEVOEEPEFVLLPQRLSIDNSKKGESOLRYROSLRPDHG 15704
QY 21 ----- 20
Db 15705 LYMIKENDHIAKAPCVSVSLDTPGPPINFPEDIRKTSVLCKWEPPLDGGSEIINTY 15764
QY 21 ----- 20
Db 15765 LEKKDKTPDSEMIYVITSLRHCKYSYKLIIEGKEYLFRVAENRFGPGCVSKPLVAK 15824
QY 21 ----- 20
Db 15825 DPGPPDAPDKPIYDVTNSMMLYKMNPKDNGSPIIGYWLIEKREVNSTHMSRVNKSILN 15884
QY 21 ----- 20
Db 15885 ALKANVDGLLEGLYVFRVCAENAGPGKFSPPSDPKTAHDPISPGPPIPRVTDTSST 15944
QY 21 ----- 20
Db 15945 IELEWEPFAFNGGEIVGYFDKQLVGTNMSRCKTKMKIVQYTVKEIREGADYKLRVS 16004
QY 21 ----- 20
Db 16005 AVNAAGEPFGTQPVYAEQEPRAVELDVSYKGIOIMAGKTLRIPAVVTGPRVPTKV 16064
QY 21 ----- 20
Db 16065 WTKEGELDKDQVVIDNVGKSELITIKDLRKHGRVYTATNCGSKFAARAVEVDVP 16124
QY 21 ----- 20
Db 16125 GPVLDKPYVTNRKCLLWMSDPEDDGSEITGFIIEKDAKMTWROPITERSKCDIT 16184
QY 21 ----- 20
Db 16185 GLLEGEYKFRVIAKNGKCGGPPVEIGPLAVDLPSPSPERLYTERTSTIILDKKE 16244
QY 21 ----- 20
Db 16245 PRSNGSGPIQGYIIEKRHRDKDPERVNRKLCPTTSFLVENLDEHOMYEFVKAVNEIGE 16304
QY 21 ----- 20
Db 16305 SEPSLPLNVYIODDEVPTIKRLSVRGDTIKVKAAGEVHIIPADVTGLPMKIEKSKNET 16364
QY 21 ----- 20
Db 16365 VIEKPTDALQITKEEVSRSSEAKTELSPKAVREDKGYVTVFNRLGSVFRNVHVEYDR 16424
QY 21 ----- 20
Db 16425 PSEPPNLAVTIDIKAESCYLTWADPLDNGSEITHYVIDKRDSRKAEMEYTNATVEKR 16484
QY 21 ----- 20
Db 16485 YGIMWKLIPNGOYEFVRVAVNKKYISDECKSDKVVIODPYRLPGPGKPLVARTKSMUY 16544
QY 21 ----- 20
Db 16545 SWTPLDNGSGPIITGYWLEKREBGSYWSRVSRAPITTKVGLKGVFNPRLLIEGVKQFR 16604
QY 21 ----- 20
Db 16605 AMAINAAGIGPSESPDEYVADPIFPGPSPCEYKDKTKSSISLGKMPRAKDGSPIK 16664
QY 21 ----- 20
Db 16665 GYIVEMQEBGTTDMKRVNNEPKLITTCQVVPNLKELRKRYRVRKAVNAGESESPDITG 16724

QY 21 ----- 20
Db 16725 EIPATDIEEPEVETIDIGADOLCKAGSOIRIPAVIKRPTPKRSWEFDGAKKAMDG 16784
QY 21 ----- 20
Db 16785 VHDIPEDAOLEAENSIVIIIECKRSHHTGKYSITAKKAGOKTANCRCRVKMDVPGPKD 16844
QY 21 ----- 20
Db 16845 LKVSIDTRGSCRLSWKMPDDGGRIGYIEKRTIDGKAMTKVNDGSGTTFVPPDLS 16904
QY 21 ----- 20
Db 16905 EQQFFVRVRAENRFGIGPPVETIQRTARDPYPPDPPIKLKIGLITKNTVHLSMKPPKN 16964
QY 21 ----- 20
Db 16965 DGGSPVTHYIECLAMPDTGKKRAMQCNKRDVELOFTYEDLVEGGEYEFVKAVNAA 17024
QY 21 ----- 20
Db 17025 GVSRRSATVGPCDCORPDMPSIDLKEMEVEEGTNVNIYAKIKGVPEPPLTWFKAPPK 17084
QY 21 ----- 20
Db 17085 PDNKEPVLIDTHVKKLVYDDCTCTYIFQSRSDTGLYTTAVNNLGTASKEMLNVLGRP 17144
QY 21 ----- 20
Db 17145 GPVGPPIKFESVADQMTLSWFPKDDGSKITNVIYEKREANKTKVHVSSPEKCTYT 17204
QY 21 ----- 20
Db 17205 IPKLLBHEHYFRIMAONKYIGEPIDSEETARNLFSVGADPKPTVSVTRNSMTVM 17264
QY 21 ----- 20
Db 17265 EEPYDGGSPYTGWLEMKDITTSKRMKRVNRDPKAMTIGVSYKVGLIGSDYQRRVYA 17324
QY 21 ----- 20
Db 17325 INAGVPSLPSDPATARDPIAPGPPPKVDWTSSADLEMSPLKDGSKVGIY 17384
QY 21 ----- 20
Db 17385 EYKEBKEEWEKGDKEVRGTLVYGLKEGAFYKFRVSAVNIAGIGEPEVTVIEMKD 17444
QY 21 ----- 20
Db 17445 RLVSFDLQDASVDRIVHAGVIRIYVSGKPPPTVTWMNNERTLQOEATIEETALS 17504
QY 21 ----- 20
Db 17505 SSMVINKQSHOGVYSLAKNAGEKRTIIVDVLDPGVGTPLAHNLJNESCKLW 17564
QY 21 ----- 20
Db 17565 FSPEDDGGSPITNVIYEKRESDRAMPVYIYTRONATVQGLIGKAVFRIAENSIG 17624
QY 21 ----- 20
Db 17625 MGFVETSEALVIREPITVPERDELEVKEVTKNTVLTLMNPPKYDGSSEIIINYVLESRL 17684
QY 21 ----- 20
Db 17685 IGTEKHAKYTNNDLSRKTYVGLKEGDIYEVVSAVNIYGGKPSCTKPIYCKDELAP 17744
QY 21 ----- 20
Db 17745 PTLHDFRDLKLTIRGEAFALIGRSGKPKPVSWFKDEADVLEDRTIKITTPATLLE 17804

QY 21 ----- 20
Db 17805 KIKAKRSDSGKYCVYENSTGSRKGCQVNVNDRPGPPVPSFDEVTIDYINYSKPP 17864
QY 21 ----- 20
Db 17865 DDGSKITNVIIEKKEVKDVMMPVTSASAKTTCVSKLLEGKDIYFRHAENLYGSDP 17924
QY 21 ----- 20
Db 17925 LVSDSKAKDRFRVDPADQPIVTEYTKDSALVTMKNPHDGGKPIINYIIEKRETSKRW 17984
QY 21 ----- 20
Db 17985 ARVTKDPIHPYTKFRVPDLLESCQYEFVSAENEIGIDPSPSKRVFAKDIAPSPPV 18044
QY 21 ----- 20
Db 18045 NPEALDTCNSVDLTWQPPRHDSKILGYIEYQKVGDEWRANHTPESCPETKYKYT 18104
QY 21 ----- 20
Db 18105 GLRDGQTYKFRVLAVNAAGESDPAHYPEVLYKDLPELILIDANMARQHIKYGDTLR 18164
QY 21 ----- 20
Db 18165 LSAIINGVPEPKVYTWKEDRDAPTKARIDVTPVGSLEIRNAHEDGIIYSLTVENPAGS 18224
QY 21 ----- 20
Db 18225 KIVSVAVLVLDKPGPPRDLVESEIRKDSCTLWKPELDGGSVITNYYVERDVAQAOMS 18284
QY 21 ----- 20
Db 18285 PLSATSKKSHFAKHLNEGNOYLFRVAENOYGRGPVETPKIYALDPLHPGPKDLH 18344
QY 21 ----- 20
Db 18345 HVDVKTESLVNMRPDRDGSPIYGLVYQEGSTODWIKFKVTNLECVYGLQOGKT 18404
QY 21 ----- 20
Db 18405 YRFRVAENIYVGLGLPDTTIPLECOEKLVPSEVLDVKLIBGLVYKAGTVFRPAIRGV 18464
QY 21 ----- 20
Db 18465 PVPYAKWTTDGESEIKTDEHYIYETDNFSSVLIKNCLRDGTGEYQITVSNAAGSTYAVH 18524
QY 21 ----- 20
Db 18525 LTVLDPGPPGTINILDPTEPHMTISWQPKDDGSPVINYIEKQDTRDKTGWVSSG 18584
QY 21 ----- 20
Db 18585 SSKTKLKIPIHLOKGEHYFRVRAENKIGVGPPLDSTPTVAAHKSPSPSGKPVVTDITE 18644
QY 21 ----- 20
Db 18645 NAAVSWTLPKSDGSPITGYMERREVTGKWRVYNTPIADLKFRYGLYEGNTYERV 18704
QY 21 ----- 20
Db 18705 FAENLAGLSKSPSSDPKACRPKIPGPPINPKLKDSRETAJLWTKPLSDGSPILG 18764
QY 21 ----- 20
Db 18765 YVVECOKPGTAQWNRINKDELIRQCAFVVPGLISGNEYRFRKAKANIYGESEPRELASV 18824
QY 21 ----- 20
Db 18825 IAKDILHPEVELDVTCHDVITVRVGQIRILARVKGHPEBDITWTEGKVLVREKRVDL 18884
QY 21 ----- 20

Db 18885 IODLPRVELQIKKAVRADHGRKIIISAKNSGHAQSAIYVLDPRGCGNLAKYNTNTKEN 18944
 QY 21 ----- 20
 Db 18945 CTISWENPLDNGSGEITNFIEYERKPNOKGMSIVASDVTKLIKANILANNEYFRVCAE 19004
 QY 21 ----- 20
 Db 19005 NKVGVGPTIETKPIILAINPIDRGEPEHLIADKCKTVEYILKWRDDYDGGSPNLSTYH 19064
 QY 21 ----- 20
 Db 19065 ERLKGSDDWERYHAKSIEKETHYVDCVENQIYEFRTKNGESDVKTEEVYKED 19124
 QY 21 ----- 20
 Db 19125 LQKPVLDLKLISVLTVRAGDTIRLEAGVRGKPPPEVAMTKDKATDLTRSPVKIDTRAD 19184
 QY 21 ----- 20
 Db 19185 SSKFSLTKAKRSGDKYVVTATNTAGSFVAIYVNLDPKGPVBNLKIYVSSDRCYCW 19244
 QY 21 ----- 20
 Db 19245 DPEDDGGCEIQNTILEKCEKRWVSTYSATVLTGTTVRLIEGNEYIFRVAENKIG 19304
 QY 21 ----- 20
 Db 19305 TGPTESKPVIAKTKYDKPGHRDPEVTKVKSEMTVWNPPEYDGGKITGYLEKKEK 19364
 QY 21 ----- 20
 Db 19365 HSTKRWVPNKSAIEBRKAKVONLLPDHEYOFRVAENIGIGESLSPRVAAKDIERP 19424
 QY 21 ----- 20
 Db 19425 GPTNFRVDTTKHSITLGMGKPYDGGAPLIGYVEMRPRIADASPEGKRCNAAOL 19484
 QY 21 ----- 20
 Db 19485 VRKEFTVSLDENOEYERVCANOQVIGRAELKEAIKREILPEPIDLDSMRKLYI 19544
 QY 21 ----- 20
 Db 19545 VRAGCPILFALVGRPARPVTKWRKVGIDNVRRGOVDLVDTMAFLVLPNSTRDGKYS 19604
 QY 21 ----- 20
 Db 19605 LTLVPAGEKAVFVNVRLDTPGPVSDLKVSVDTKTSCHSWAPENDGGSQVTHYIEK 19664
 QY 21 ----- 20
 Db 19665 READRKSTVTPPEYKTSFHYTNLVPGNEYFRVTAVNEGPGVPIDVPRVLAUSDPLS 19724
 QY 21 ----- 20
 Db 19725 EPDPRLKEVTEMTKNSATLAMPRLRDGAKIDGYITSREBOPADRWTEYSVVDLS 19784
 QY 21 ----- 20
 Db 19785 LVVTGLKEGKRYKFRVARNVAVSLPREAGVYEAKEQLLPKILMEPOITIKAGKLR 19844
 QY 21 ----- 20
 Db 19845 IEAHYGRPHPCKKKKEDEVYTSHLAVHAKDSSILLIKDVTNRDGSYYSLAENS 19904
 QY 21 ----- 20
 Db 19905 GMDTQIKIVVMDAPGPDPDISDIDADACSLWHIPLDGGSNITNYIWEKCDVSRG 19964
 QY 21 ----- 20

Db 19965 DWVTALASVTKTSCHVKLIPGOEYIFRVAENRFGISEPLTSPKVAOPEFVSPKPN 20024
 QY 21 ----- 20
 Db 20025 ANVTYVNRKDCIFVANDRPDSGGSPYIGYLLERKERNSLWVANKDILVRSTETPCAGLV 20084
 QY 21 ----- 20
 Db 20085 EBLEYSFRIYALNKAQSSPPSKPTETVARMVPDPGKPEVIDYTKSTVSLIMARPKHDG 20144
 QY 21 ----- 20
 Db 20145 GSKIIGYVEACKLPDGRKWRNCNTAPHQIPQEYTAGLEKKAQYQFARLARTAVNISPP 20204
 QY 21 ----- 20
 Db 20205 SEPSDPVTILAEVNPRIIDLVSAMKSLTYVAGTVCLDATVFGKPMPTYSMKKDGLTK 20264
 QY 21 ----- 20
 Db 20265 PABGIKMAQRLCTLLEFSVNRKDSGYTTIAENSSGSKSATIKLVLDKPGPASVKI 20324
 QY 21 ----- 20
 Db 20325 NKYSDRAMLSWEPLEDGSEITNYIVDKRETSRPNMAQVSAVPIITSCSVEKLIBGHE 20384
 QY 21 ----- 20
 Db 20385 YQFRICAENKYGVDPVTEFPAIAKNPYPGRCDPYISNTKDHMTVSMKPRADGGS 20444
 QY 21 ----- 20
 Db 20445 PITGYLLEKRETOAVNMTKVNRRPIIERTLKATGLQEGTEYFRVTAINKAGPKPSDAS 20504
 QY 21 ----- 20
 Db 20505 KAYVARDPOYPAPPAFPEYVYTTSSVSLSMGKPAIDGSGPIIGYLYEVKRAQSDMNV 20564
 QY 21 ----- 20
 Db 20565 CNLPQNLQKTRFVINGLMDTQYQFRVYAVNKIGYSDPSVDPKHYPKDILIPPEGELDA 20624
 QY 21 ----- 20
 Db 20625 DLKRTLLRAGVTMLYVVKGRPPPKITWSKPNVNLDRIGLIDKSTDFTELCENVN 20684
 QY 21 ----- 20
 Db 20685 KYDAGKIITLLENSGKRETIYVAVYLDTPGPVNVYVKEISKDSAYTWEPPIIDGGSP 20744
 QY 21 ----- 20
 Db 20745 IINYVQKRDAERKSWSTVTECSKTSFRVANLEBGSYFFRVAENEXIGDPGETRDA 20804
 QY 21 ----- 20
 Db 20805 VKASQTPGPVVDLKVRVSCKSCSIGMKPBSDGSRILIGYVVDLLEENKQORWMSLS 20864
 QY 21 ----- 20
 Db 20865 LOYSANDLIEGKREYTFRVSANENEGTSPSEITVVARDDVAPBDLKLGLPDLCLAKEN 20924
 QY 21 ----- 20
 Db 20925 SNFRLKIPIKGKAPSVSMKGEDPLATDRVSVSSAVNTLIIYDCQKSDAGRYTTL 20984
 QY 21 ----- 20
 Db 20985 KNAVGTKEGTISTIKVVGKCPITGPILKDEVTAEANTLWAPPKDGGSEITNYILKRD 21044
 QY 21 ----- 20
 Db 21045 SVNKKWTCASAVOKTTFRTVRLHEGMEYTFRVSANENKYGVEGLKSEPIVARHPEDVPD 21104

```

QY 21 ----- 20
Db 21105 APPPNIVDRHDSVLTWTDPKTKGSPITGYHLEFKERNSLMKRANKTPIRMRDPKV 21164
QY 21 ----- 20
Db 21165 TGLEGEYEFRRWAINLAGVKSLSPEPVVALDIPDPGKPEVINITRNSVTLLIMTEP 21224
QY 21 ----- 20
Db 21225 KYDGHKLTYIVEKRLPSKSMKANHVNEBACFTYTDLVGGKYEFRIRAKNTAGAI 21284
QY 21 ----- 20
Db 21285 SAPSESTETICKDEYEAPITVLDPTIKDGLTIRAGDTIVLNAISILGKPLPKSSMSKAG 21344
QY 21 ----- 20
Db 21345 KDIRPSDITQITSTPTSSMLIKYATRKRDAGETTITATNPGTVEHVKYTVLDVPGPPG 21404
QY 21 ----- 20
Db 21405 PVEISNVAERKATLTWTPLEDGSPIKSYILERETSRLMTVVSIEDIOSCRHATKLI 21464
QY 21 ----- 20
Db 21465 QGNEYIRVSAVNHYGGEVQSEPVKMDRFGPPPEKPEVSNVTKNATVSMKRPVD 21524
QY 21 ----- 20
Db 21525 DGGSEITGYHERREKKSLSRNVRAIKTPVSDLRCKVTGLQEGSTYEFVSAENRAGIGPP 21584
QY 21 ----- 20
Db 21585 SEASDVLMKDAAYPGPPSPNHTDPTTKKSASLAWGKPHYDGLBITGYVEHOKVDE 21644
QY 21 ----- 20
Db 21645 AMIDDTGTALRTQFVVPDLQTEKYNFRIASINDAGVEPAVDPVEIVEREMADFE 21704
QY 21 ----- 20
Db 21705 LDAELRTLVVRAGLSIRIFVPKGRAPAEVWTKONINLKNRANIENTESTILLIPEC 21764
QY 21 ----- 20
Db 21765 NRYDTGKFWMTIENPAGKSGFVNVRLDTPGPVNLNRPDITKDSVTLHMDPLIDGGS 21824
QY 21 ----- 20
Db 21825 RITNYIVEKREATRKSSTATTKCHKCTYVYGLSEGCYFFRVMAENEXYIGEPTETWE 21884
QY 21 ----- 20
Db 21885 PVKASEAPSPDLSINMDITKSTVSLAMPKPKHDGSKITGYVIEAQKSGSDQWHTHTTY 21944
QY 21 ----- 25
Db 21945 KGLECVVRNLTGEBEYTFQVMAVNSAGRSAPRESRPVYKEQTMLEPDLRGTYOKLVIA 22004
QY 26 ----- 25
Db 22005 KAGDNIKVEIPVLGRPKPTVTKKGOILKQTORVNEFTTATSTIININECVASDSGPPY 22064
QY 26 ----- 25
Db 22065 LTAARNIVEGVDTITQVHDIPGPTGPIKFEDEVSSDEVTFMSDPENDGVPISNVYE 22124
QY 26 ----- 25
Db 22125 MROTDSTTWELATVYIRTTYKATRLTGLEIYQFVKAONRNVGVPGITSACIVANYPFK 22184

```

```

QY 26 ----- 25
Db 22185 VPGPGHPQVTAATKDSMTISWHEPLSDGSPILGYHERKENGLIMQVSKALVPGNI 22244
QY 26 ----- 25
Db 22245 FKSSGLTDIAYEFVIAENMAGSKRSPSEMLALDIPDPGKRPVPLNTRHTVTLKN 22304
QY 26 ----- 25
Db 22305 AKPEYTGCFITISYIVEKRDLPNGRMKANFSNILENEFTVSGLEDAAYEFVIAKNNA 22364
QY 26 ----- 25
Db 22365 GAISPPSPSDAITCRDVEAPKIKVYKFKDVIILKAGEARLEADVSGRPPIWENSK 22424
QY 26 ----- 25
Db 22425 DGELEGTAKLEIKIADFSTNLVNRDSTRDSGAYTLTATNPGFAKHIFNVKVLDRPGP 22484
QY 26 ----- 25
Db 22485 PBGPVATVEVTSBKCVLSNFPPLDGGAKIDHYIYOKRETSRLAMTNVASEVQVTKLVT 22544
QY 26 ----- 25
Db 22545 KILKNEYIFRVMAVNRKYGVEPLESEPVIAVNPYGPDPKPNPEVTTITKDSNVVCGH 22604
QY 26 ----- 25
Db 22605 PSDGSEIINIVERRDAGORWIKCNKTYLTDLRKVSGLTEGHEVEFRIMAENAGI 22664
QY 26 ----- 25
Db 22665 SAPSPSPFYKADIFYKRPGRPNRVLDTSRSSISIAMNKPIYDGGSEITGYVETALP 22724
QY 26 ----- 25
Db 22725 EEDMOIVTPPAGLKATSYITIGLTENOEYKIRIYAMNSEGLGEPALVGPCKADRMIP 22784
QY 26 ----- 25
Db 22785 PEIELDADLRKYVTTIRACCTLRLEVPKGRAPAEVKMADHGESLDKASISTSSYTLII 22844
QY 26 ----- 25
Db 22845 VGNVNRFDGKYLIVENSSGSKSAFVNVRLDTPGPPDLKVEVTKTSVTLTMDPPL 22904
QY 26 ----- 25
Db 22905 DGGSKIKNYIVEKRSTRKASTVATNGHKTSMKVDQLQEGSYFVRLAENEGIGLPA 22964
QY 26 ----- 25
Db 22965 ETAEVYKASERLPPGKITLMDVTRNSVSLSWKEPHDGSRIIGYIVEMOTKSGDKWAT 23024
QY 26 ----- 25
Db 23025 CATVWTEATITGLQGEBSYSPVSAONEKGISDPRQLSVPVIAKNDVIIPAFKILTEWTF 23084
QY 26 ----- 25
Db 23085 TVLAGEDLKVDVPFGRPTPAVTHKDNVPLKQTVRNAESTENNSLITIRDACREDVGH 23144
QY 26 ----- 25
Db 23145 YVVKLTNSGAEIETLVNIYVLDKPGPTGPVYKMADEVTAISITLSKGPYKDGSSINNYI 23204
QY 26 ----- 25
Db 23205 VEKRDSTTWQIVASATVARTTIKACRLKTCGEVQFRIAENRNGKSNVLSNPTVAQYP 23264
QY 26 ----- 25

```

```

Db 23265 FKVPGPCTPVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERNSLMWKLNKTIPIQ 23324
QY 26 ----- 25
Db 23325 TKFKTGLEBGVEYEFVSAENITVGIGKPSKSECVYARPCDPPGRPEALIVTRNSVTL 23384
QY 26 ----- 25
Db 23385 QWKKPYDGGSKITGYIVERKELEPEGHMKASFTNLIIDTFEVTGLVEDHREYFVIA 23444
QY 26 ----- 25
Db 23445 AAGVFSESESTGAIARDEVDPPRISMDPKYDTIVYHGESKVDADYKPIPTIQ 23504
QY 26 ----- 25
Db 23505 IKGOELSNFARLEIKSTDFATISLVADAVRDSGNILKAKNAGERSVTYVNVKVLDRP 23564
QY 26 ----- 25
Db 23565 GPPGPVVISGVTAEKCTLANKPPLDGGSDIINTYVERRETSRLVWTVVDANVOTISCK 23624
QY 26 ----- 25
Db 23625 VTKLLEGNEYFRIMAVNKYGVGEPLSEPVYAKNPFVDPAPKPEVTVTKDSMIYVW 23684
QY 26 ----- 25
Db 23685 ERPA5DGGSEILGYLEKREKIRMTCHKRLIGELRLVTLGIENHDEYFVSAENAA 23744
QY 26 ----- 25
Db 23745 GLSESPPSAOKACDPIYKKGPPNNKRVIDITRSSVFLMSKRIYDGGCEIGIYIEKC 23804
QY 26 ----- 25
Db 23805 DVSQWMTCTPTGINKTINIEVEKLEKHEYNFRICAINKAGVGHADVPPIVEKL 23864
QY 26 ----- 25
Db 23865 EAPDIDLLEKLIINIRAGSLRLFPYIKGRPTPEYKMGVNDGEINDAIIIDVTSFTS 23924
QY 26 ----- 25
Db 23925 LVLDNVNRYDSGKYTLTLENSGTSKSAFVTVRVLDPSPVNLKVTETIKDSVSIWMBP 23984
QY 26 ----- 25
Db 23985 LLDGSKIKNTYVERRETRKSYAAVVTNCHKNSWKIDLOEGSYFFRYTAENEYIGIL 24044
QY 26 ----- 25
Db 24045 PAQTAADPIKVAEVPQPKITVDVTRNSVLSMTKPEHDGSKIIQIYIVEMQAKHSEKW 24104
QY 26 ----- 25
Db 24105 SECARVKSLOAVTNLTQGEELYFRVAVNNEKGRSDPSRLAVPIYAKDLVIEPDKAFS 24164
QY 26 ----- 25
Db 24165 SYSVOVGODLKEIYVPSRPPKPTITWKDGLPKQTRINVTDSLDLTLSIKETHKDDG 24224
QY 26 ----- 25
Db 24225 GQYGITVANVVGOKTASTIEITWLDKDPKPVKRDVSAESITLSMNPPLYTGCCQITN 24284
QY 26 ----- 25
Db 24285 YIVQRDPTTTTVMVDVSAFVATTLKVTKLKTGEYQFRIFAENRYGOSFALSDPIYAQ 24344
QY 26 ----- 25

```

```

Db 24345 YPKKEPGPPGFEPATAISKDSNVIQWHEPVNNGSPVIGYHLERKERNSLMTKVNKTI 24404
QY 26 ----- 25
Db 24405 HDTOFKAONLEBIEYEFERYVAENITGVGKASKNSECVYARPCDPPGTPPIWKNKEI 24464
QY 26 ----- 25
Db 24465 TIQWTKPYDGGSMITGYIVERKDLPDGRMKKASTNVIETQFVYSGLTEDQRYEFVIA 24524
QY 26 ----- 25
Db 24525 KNAAGAIKSPSDSTGPIYAKDEVELPRISMDKFRDTIVNAGETFRLEADVHGKPLPTI 24584
QY 26 ----- 25
Db 24585 EWLKGDKEIESARCEIKNTDFKALLIYKDAIRIDGQOTILASNVAGSKSEPVNVKVL 24644
QY 26 ----- 25
Db 24645 RPPPGPVQVYGVYSEKSLTWSPPLDGGSDISHYVERKETSRLAMTVYASEVVTNS 24704
QY 26 ----- 25
Db 24705 LKVTKLLEGNEYFRIMAVNKYGVGEPLSEAPVLMKNPFVLJGPPKSLVTNIAKDSMTV 24764
QY 26 ----- 25
Db 24765 CWNRPDGGSEILGYIVERKDRSGIRWIKCNKRRTIDLRRLVGLTDEHDEYFVSAEN 24824
QY 26 ----- 25
Db 24825 AAGVSESPATVYKACDVPFKPPTNAHIVDTTKNSITILAMKPIYDGGSEILGYVE 24884
QY 26 ----- 25
Db 24885 ICKADEEMOQIYPTOTGLRVTREISKLTLEHOEYKIRVCAIANKVGLGATSVPYKPED 24944
QY 26 ----- 25
Db 24945 KLEAPELDSELKGIIVRAGSARIHIPFKGRPTPEITWSEGEFTDKQIEKGVNY 25004
QY 26 ----- 25
Db 25005 TOLSDICRNDAGKYLKLENSGSKSAFVTVKVLDPGPONLAVNEVRKDSAPLWE 25064
QY 26 ----- 25
Db 25065 PPIIDGAKYKNVIDKRESTRKAYANVSSKSKTSFKYENLTGALYFRVMAENERGV 25124
QY 26 ----- 25
Db 25125 GVPETVDAVKAEPSPPGKVTLLDVQSASLMEKPEHDGSRVIGYVEMQPKGTE 25184
QY 26 ----- 25
Db 25185 KMSIYAESKYCAVAVTGLSSGOEYQFRVAKYNEKGS DPRVLGVPYIAKDLTIOPSILKP 25244
QY 26 ----- 25
Db 25245 FNTYSTOAGEDLKIEIPYIGRPRNISWYKDEBPLKQTRVNVETATSTVLHIKGNKD 25304
QY 26 ----- 25
Db 25305 DFGKYVTATNSAGTATENLSVILEKPGPVGPVRFDEVSADEVVLSWEPAYTGGOI 25364
QY 26 ----- 25
Db 25365 SNTIYERKDTTTTWHMVAFTVAKTIKITKLKTGEYQFRIFAENRYKSAPLDSKAVI 25424
QY 26 ----- 25
Db 25425 VQYFKKEPGPPGTPFVTSISKQMLVQWHEPVNDGGTKIIGYHLQEKERNSLMWKLNKTI 25484

```

QY 26 ----- 25
Db 25485 P10DTKRTTGLDEGLEFEKVSANENIVGIGKESVSECFVARDPCDPGRPAIVITRN 25544
QY 26 ----- 25
Db 25545 NVTLMKKAPAYDGSKITGTYIVEKKDLPGGRMKASFTNVLETFVSGLVEDQRYEFV 25604
QY 26 ----- 25
Db 25605 IARNAAGNFSEPSDSCATARDEIDAPNASLDPKKYIVAHAGTFVLEADIRGKPIR 25664
QY 26 ----- 25
Db 25665 DVVMSKDGKELEETAARMEIKSTIOKTTLVVKDCIRTDGQYILKLSNVTGTSIPITVK 25724
QY 26 ----- 25
Db 25725 VLDRPGPBGPLKVTGVTAKCYLANNPLODGGANISHYIIEKRETSRLSWTQVSTEVO 25784
QY 26 ----- 25
Db 25785 ALNFKVTKLPGNEYIFRVAHVAKYIGEPLESGPYTACNPKYKPGPSPPEVSATTKDS 25844
QY 26 ----- 25
Db 25845 MVTMARPYDDGGTEIEGYILEKRDKEGVKWKCNKKTILDLRLVTLGTEGSHSEFRVA 25904
QY 26 ----- 25
Db 25905 AENAGVGPSEPSVFRACDALYPPGPPSNKVTDTSSSVSLANSKPIYOGGAPVKGY 25964
QY 26 ----- 25
Db 25965 VVEKKAADWETTCYPTGLOGKOPTVTKLKNTEYENFRICAINSEVGEPATLPGSVV 26024
QY 26 ----- 25
Db 26025 AQERIPPELELDLRLKVVVLRASATLRLFYTIKGRPEVKWEKAGILDRAOIEVT 26084
QY 26 ----- 25
Db 26085 SSFTMLVINDVTRFDSGRNLTLENNSSGSKTAFVNVRLVDSAPVNLIREVKDVSYL 26144
QY 26 ----- 25
Db 26145 SWEPLIDGAKITNYIVEKRETRKAVATITNCTKTFRIENLOEGCSYFVRLASNE 26204
QY 26 ----- 25
Db 26205 YGIGLPAETTEPVKSEPLPRGRVTLVDYTNNTATIKWEKESDGSKITGYVVMQTK 26264
QY 26 ----- 25
Db 26265 GSEKMSCTQVKTLEATISGLTAGEEYVPRVAVNEKGRSDRQLGVPIANDIEIKPSV 26324
QY 26 ----- 25
Db 26325 ELPEHTFNKAREOLKIDVPFKGPOATVNMKRDGOTLKETTRVNVSSKVTSLSIRKA 26384
QY 26 ----- 25
Db 26385 SKEDVGYELCVSNSAGSITVPTITIIVLDRPGPIRIDEVSCHSITISNPPEDGCG 26444
QY 26 ----- 25
Db 26445 QISNIVKEKETTSTWHIVSOAVARTSIKIVRLTTGSEYOPRCAENRYGKSSSESA 26504
QY 26 ----- 25
Db 26505 VVAEYFSPSPGPGTIPKVVHATKSTMLVTWQVPVNDGSRVIGYHLEKENSILMSKAN 26564

QY 26 ----- 25
Db 26565 KILIDTQMKVSGLDEGLMEYERYVAENIAGIKSKSCPEVPARDPCDPGQPEVTNIT 26624
QY 26 ----- 25
Db 26625 RKSVALKMSKPHYDGAKITGYIVERRELDGRMLKCNVTNIOETTFEVTLETDQRYEF 26684
QY 26 ----- 25
Db 26685 RYFARNADSVSEPSSTGPIIVKDDVEPPRVMDYKFRDVIYVAKAGEVLKINADIAGR 26744
QY 26 ----- 25
Db 26745 LPVISMAKDGIIEERAREIISTDNHTLLVVKDCIRRDTOGYVILTKNAGTSVAVNC 26804
QY 26 ----- 25
Db 26805 KVLDRPGPAGPLEINGLTAEKCSLSMGRPOEDGADIDYIYVERKRETSHLAWTICEGEL 26864
QY 26 ----- 25
Db 26865 QMTSCKVTKLKGNFYIFRVTVKVKYGVGEPLSVATKALDPFTVPSPISLEITSVKE 26924
QY 26 ----- 25
Db 26925 SMTLCMRPESDGSSEISGYIIEERKNSLRVNRNKKPVYDLKVKSTGLREGCEYERY 26984
QY 26 ----- 25
Db 26985 YAEANAGLSLESETSPILIRADPVFLPSPSPKPIVDGKTTITIAWVKPLFDGADITG 27044
QY 26 ----- 25
Db 27045 YVEYKKSDDIDWKTISIOSLAGTEYISGLTGAIFYFRKVSANKVAGSDPSDSDPOIA 27104
QY 26 ----- 25
Db 27105 KEREPEPLDIDSEMRKTLIVKAGASFTMTVPFGRPVPNVLSKPPDLRTRAYDITD 27164
QY 26 ----- 25
Db 27165 SRTSLTIENANRNDGKTYTLIQVLSAASLTIVVKVLDPPGPTNITVDVTESAVLS 27224
QY 26 ----- 25
Db 27225 MDVPENDGAPVKNYHIEKREASKKAWSVTNNGCNLSYKVTNLQEGAIYFFRYSGENEF 27284
QY 26 ----- 25
Db 27285 GVGIPAEKTEGVKITKPSPEKLGVTISKDSVSLTWLKEPHDGSRIYHYVEALEKG 27344
QY 26 ----- 25
Db 27345 QKNWKCVAKSTHHVYVGLRENSEYFFRVFAENQAGLSDPRELLPLVILKEOLEPEID 27404
QY 26 ----- 25
Db 27405 MKNFESHVYVRAGSNLKVDPISGRPLPKVTLSDGVPLKATWRENTETIAENLTINLK 27464
QY 26 ----- 25
Db 27465 ESVTADAGREITTAANSSTGTFKAFINIVLDRPGPTGPVVISDITEESVTLKWEPPKYD 27524
QY 26 ----- 25
Db 27525 GGSQVNTVILKRETSAAVTEVSATVARTMKVMKLTIGEEYOFRIKABNRGISDHD 27584
QY 26 ----- 25
Db 27585 SACYVRLPYTTPGPSPSTPMVNTVNTRESITVGWHEPVSNGSANVGYHLEMKDRNSILMQ 27644
QY 26 ----- 25

Db 27645 KANKLVIRTHHEKVTISAGLIYEEFVNAENAGVGRPSHPSEPVLAIDACEPRNRVIT 27704
QY 26 ----- 25
Db 27705 DISKNSVLSMWOQAPDGGSKITGYIERRDLPDGRMTASFTNVTETOTIISGLTQNSQ 27764
QY 26 ----- 25
Db 27765 YEEFVARNAGVSNPSEVGPITCIDSYGCVIDLPLEYEVVYRAGTSVKLRAGIS 27824
QY 26 ----- 25
Db 27825 GKPAPIEWKDDKELQTNALVCVENTDILIKDADRLNSCYELKLRNMGASAT 27884
QY 26 ----- 25
Db 27885 IRVQJILDKPGPGPIEFKIVTAEKITLMLRPDADGAKITHYIEKRETSRVMSVS 27944
QY 26 ----- 25
Db 27945 EHLBECITTTIKIGNEIYFRVAVNKGICIGPLSDSVAKNAFVTPGPGIPEVTKI 28004
QY 26 ----- 25
Db 28005 TKNSMVFVNSRPADGSDISGYFLEKRDKSLGMEFKVLEKITRDTROKVTGLTENSQ 28064
QY 26 ----- 25
Db 28065 YRVCAVNAAGOPPESEPEYKADPIDPPPAKIRIADSTKSTITLGMKRPVYDGS 28124
QY 26 ----- 25
Db 28125 VTGVEIIRGEBEEMTVTSGEVRTTEYVSNLPGVNYFVSAVNAAGGPEIEMN 28184
QY 26 ----- 25
Db 28185 EPVQAKDLEAPIDLDVALRTSVIAKAGEDVOLLPEKGRPPVTWKRDEKNLGSAR 28244
QY 26 ----- 25
Db 28245 YSTENTDSSILITPOVTRNDTGKYLITLENGGEPRSSVSVKVLDTPAACOKLVKHV 28304
QY 26 ----- 25
Db 28305 SRGTVLLMDPPLIDGSPILINVIKRDATKRTWSVSHKCSSTSEKILIDSEKTPFE 28364
QY 26 ----- 25
Db 28365 RVIAENEIGICEPCETTEPVKAAEVPAPIDLSKDSSTKSTVILSTKPDFDGSVITTE 28424
QY 26 ----- 25
Db 28425 VVERKKGEGTWSHAGISKTCEIVSQLKEOSVLEFVFAKNEKGLSDPYTIGITVKEL 28484
QY 26 ----- 25
Db 28485 IITPEVLDSDIPGAQVTVRIGHNVHLEPYKGRKPSISWLKDLPLKESEFVRFSKTEN 28544
QY 26 ----- 25
Db 28545 KIITSLNAKKEHGKITTIVILDNAVCRIANPITVITIGPSPKRGPIRDEIDADVILS 28604
QY 26 ----- 25
Db 28605 MDVEDNGGEITCYSTIEKRETSQTNMKVCSVAFTFKVPLVKDAEQFVRAENRY 28664
QY 26 ----- 25
Db 28665 GVSQPLVSVIIVAKHQFRIQPPKPVIIYVTSDGMSLTWDAPVYDGSSEVTGFHVEKKE 28724
QY 26 ----- 25

Db 28725 RNSILMQKVNTPSPISGREYRATGLVEGLDYQFVYAENSAGLSPSPDSKFTLAVSPVD 28784
QY 26 ----- 25
Db 28785 PGTPDYIDVTRETITTLKMNPLRBDGSKIVGYSTIEKROGNBRVRCNFTDVSECOTYVG 28844
QY 26 ----- 25
Db 28845 LSPGDRYEFRIIARNNAVOTISPPSOSSGIIMTRDENVPPIVEFGPEYFDGLIISGESLR 28904
QY 26 ----- 25
Db 28905 IKALVQGRPVRYTWFKDGYEIEKRMNMETDVLGSLFVRQATRDHGRVYVEAKNAS 28964
QY 26 ----- 25
Db 28965 GSAKAEIKVQODIPGKVVGPITRTNITGEMTLMWADPINDGCAPIITHYIEKRETSRL 29024
QY 26 ----- 25
Db 29025 AMALIEDKCEASVTAIKLINGNEYQFVSAVNFVGRPLSDPVVAQIQYVDPAPGI 29084
QY 26 ----- 25
Db 29085 PEPSNITGNSITLWAPESDGSSEIQYIIEERREKSTRWVKVISKRPISETRFEKVTGL 29144
QY 26 ----- 25
Db 29145 TEGNEYEFVMAENAGVGPASGISRLIKREPVNPGPPTVAVKVTDTSKVTLSLEMSKP 29204
QY 26 ----- 25
Db 29205 VFDGMEIIGYIIECKADLDGMRKVNNAECVKTRTVTDLOAGEEYKFRVSAINGAKG 29264
QY 26 ----- 25
Db 29265 DSCVETGIIKAVDRILTAPELDIDANFKQTHVVRAGASIRLFIAYQGRPTPAVMSKPSDN 29324
QY 26 ----- 25
Db 29325 ISLRADIHHTDSFTLTVENCNRRNDAGKTYTLTVENNSSKSTFTVAVLDPPGPGITF 29384
QY 26 ----- 25
Db 29385 KDVTRGSATLMADAPLDLGARIIHHYVEKRBASRRSQVISECTROIKRVNDLACVP 29444
QY 26 ----- 25
Db 29445 YFVRSAVNEVGPEYEMPEPIVATEQPAPRRLDVYDTSKSAVLAMLPDHDGSR 29504
QY 26 ----- 25
Db 29505 TGYLLEMRQKSGDFWEAGHTRKQLTFVVERLVEKTEYEFVAKANDAGYSEBRFAFSVI 29564
QY 26 ----- 25
Db 29565 IKEPOLIPTADLTGITNOLITCKAGSPPTIDVPISGRAPKVTWLEMRILKETDRVSIT 29624
QY 26 ----- 25
Db 29625 TYKDRITLVKDSMRGDSGRYFLTEMTAGVTFSVTVVYIGRPGPVGPIEVSSVSAES 29684
QY 26 ----- 25
Db 29685 CVLSWGEPRDGGTEITNYIEKRESGTTAMQLVNSVVKQIKVTHLTKYMEYSFRVSS 29744
QY 26 ----- 25
Db 29745 ENRFGVSKPLBESAPLIAEHFPVPSPAPTRPEYHYHSANAMSTRMEEPYHDGSKIGVW 29804
QY 26 ----- 25
Db 29805 EKKERTILWKENKVPCLCECNKYVTGLVEGLEYOFTYVALNAGVSKASEASRPIMQN 29864

QY 26 ----- 25
 Db 29865 PVDAPGEVTDVTRSTVSLIWSAPAYDGSKVGYIIEKRPVSEVGDGMLCNVTIYS 29924
 QY 26 ----- 25
 Db 29925 DNEFTVVALSEGDTEFRVLAKNAGVYSKSESTGPTCRDEYAPKALDLARLHGLDV 29984
 QY 26 ----- 25
 Db 29985 TIRASDLVLDAVAGKPEPKIITWKDKEJLCEKVSLOYTGKRAFAVFKFCDRSDSGK 30044
 QY 26 ----- 25
 Db 30045 YTLTVKNASGTAVSVKVLDSBPGKLTVSRTQEKTLAMSLPQEDGAEITHYIY 30104
 QY 26 ----- 25
 Db 30105 ERRETSRLNMVIVEGECPTLSYVVTLLIKNNEYIFRRAVNKYGPGVPESEPIVANRSE 30164
 QY 26 ----- 25
 Db 30165 TIPSPPGIPEEVGTGKEHIIQWTKPESDGNELSNVLVDKREKSLMRVKNKDYVVD 30224
 QY 26 ----- 25
 Db 30225 TRLKVTSLMEGCOYQFRVTAVNAGNSEPSEASNFISCREPSYTPGPSAPRVVDTTKHS 30284
 QY 26 ----- 25
 Db 30285 ISLAWTRPMYDGGTDIVGYVLEMOEKDJDQWYRVHNTATIRNTEFTVPDLKMGOKISFRY 30344
 QY 26 ----- 25
 Db 30345 AAVNVKGMSEYSESIAIEFEVERIEIPDLDELADLKKTVIRAGASLRMLWVSVGRPPV 30404
 QY 26 ----- 25
 Db 30405 ITMSKOCIDLASRAIIDTESYSLIYDKVNRDACKYITLAEONSGKSATVLVYVDY 30464
 QY 26 ----- 25
 Db 30465 PGBCPSVKAEVSRDSVTITWEIPTIDGAPVNNYIYEKREARAFKVTTKCSTILR 30524
 QY 26 ----- 25
 Db 30525 ISGLVEGTMVYFRVLBPENIYIGPCETSDAVLSEVPLVPAKLEVVDYTKSTVLAMEK 30584
 QY 26 ----- 25
 Db 30585 PLVDGGSRLGYVLEACKAGTERMKVVLTKPVLKLEHTVYSLNEGHQYLFRIAPONKGV 30644
 QY 26 ----- 25
 Db 30645 SEPRETVAVTQDLRVLPTLIDLSMPKTIHVPAGRPELVIPJAGRPPAAMFEAGS 30704
 QY 26 ----- 25
 Db 30705 KLRESERVVEHTKVAKLTIETTRDGTGEYLELKNVTGTTSETIKVLIIDKRGPPGT 30764
 QY 26 ----- 25
 Db 30765 PIKIDEIDATSIISWPEPELDGAPLGSYVVEQDRAHHPGMLPVSESVTRSTFKFTRLT 30824
 QY 26 ----- 25
 Db 30825 EGNVEYFVAATNRFGISYLOSSEVIECRSSIRIPGPETLQIFVSRDGMTLTWPPED 30884
 QY 26 ----- 25
 Db 30885 DGSQVGTGIVERKEVRADRWVRVKNKVPVLTMRYSTGLTGLEEYEHRYTAINARGSGRP 30944

QY 26 ----- 25
 Db 30945 SRPSKPIVAMDLPIAPGKQPNRVDITRTSVLSAMSVPEDBGSKVGYLIEKQKUDH 31004
 QY 26 ----- 25
 Db 31005 EWTKCNTPPTKIREYTLTLPOGAERYFRVLACNAGGPEPAVEPVGVKTEMLEYDYE 31064
 QY 26 ----- 25
 Db 31065 LBERVOEGIFVNOGVYIRLTIPIKGPPEPICWTKREGODISKRAMIATSEHTELVKEA 31124
 QY 26 ----- 25
 Db 31125 DRGSGTYDVLLENKCGKKAIVIKVRIGSPNSPBGPLEYDDIQVRSVSRPADDG 31184
 QY 26 ----- 25
 Db 31185 ADILGYILERRVPRAAWTTIDSRVGTSLVYKGLKENVEYHFRVSAENOGFISKPLKE 31244
 QY 26 ----- 25
 Db 31245 ERYTRKTPLNPPPEPSNPPEVLDVTKSSVLSWSRPKDDGSRVTGYIIEKRTSTDKW 31304
 QY 26 ----- 25
 Db 31305 RANNTQITTTMYTVGLVDAEYQFRIIAQNDVGLSETSPASEPVYCKDPPDKPSQBEL 31364
 QY 26 ----- 25
 Db 31365 ELLISKDSVTLQWKEPBCDGGKEILGYWVEYROSQDSAMKSNKERIKQKOTIGLLE 31424
 QY 26 ----- 25
 Db 31425 ATEYEFVRFAENETLSRPRTAISIKTKLISGEAPGIRKEMKDYTKLGEAOLSCQIV 31484
 QY 26 ----- 25
 Db 31485 GRPLDIKMYRGEKELIOSRKYKMSDGRHTLTVWPEQDEGVYQIATNEVEGETS 31544
 QY 26 ----- 25
 Db 31545 SKLLQATPOHPGYPLKEKYGAVSTLRHVMYIGRPAMTWFGOKLLQNSENITI 31604
 QY 26 ----- 25
 Db 31605 ENTEHYTHLVKKNVQKTHAGKYKQVLSNFGYDAILDVEIQDKPDKPTGPIVIEALLK 31664
 QY 26 ----- 25
 Db 31665 NSAVISMRPPADGGSWITNYVVEKCEAKEGAEQOLVSSAISVTTCTRLVLTENAGYFR 31724
 QY 26 ----- 25
 Db 31725 VSAONTFGISDPLEVSSVYIISPEKPGAPGKPTITAVTKDCVAMKPPASDGAKIR 31784
 QY 26 ----- 25
 Db 31785 NYLEKREKONKMLSVTTEIRETVSVKNLIGLEEFVKCENLGSEMEISEPT 31844
 QY 26 ----- 25
 Db 31845 TPKSDVPIQADHFEELNLNVRYOSNATLVCKVYGHKPIVKNYROGKEIIADGLKRI 31904
 QY 26 ----- 25
 Db 31905 QEFKGYHQLIASVTDADATVYQVRATNOGGSVGTASLEVEVPAKIHLPKLLEGKAV 31964
 QY 26 ----- 25
 Db 31965 HALRGEVVISIKIPSGRPDPVITWQKODLIDNNGHYOIVYTRFTSLVPNGVERKDAG 32024
 QY 26 ----- 25

Db 32025 FYVYCAKNNRGIDOKTVELDVADVPDPBGVAVSVDSRDSVNLWTTERPASDGSKITNTI 32084
QY 26 ----- 25
Db 32085 VEKCATTAERHLNVOGARRETRYVIMLFGKTSYQFRVIAENKFGSKPSPSEPTITKED 32144
QY 26 ----- 25
Db 32145 KTRAMNYDEVEDTREVSMTKASHSTKELYEKYMTAEDLGCEGFIHRCVETSSKTY 32204
QY 26 ----- 25
Db 32205 MAKFYVKGTDYLVKKEISILNIAHRNLIHLHESFESMEELVMTFEFISGLDIPERIN 32264
QY 26 ----- 25
Db 32265 TSAFELNEREIVSYHVCBALQFLSHNIGHDIPRENIYQTRRSSTIKIIEFGQARQ 32324
QY 26 ----- 25
Db 32325 LKQGNFRLFLTAPEYIAPRHQHDVSTATDMWSLGLTVYVLLSGINPLAETNOIIE 32384
QY 26 ----- 25
Db 32385 NIMNAEYTFDEALFKEISIPAMDVDRLLVKEKRSMTASEALQHPMLKOKIERVSTKYI 32444
QY 26 ----- 25
Db 32445 RTLKHHRYHTLLIKDLNMVVSARISCGAISRQGSVAKVAVASIEIGPVSGOIMHA 32504
QY 26 ----- 25
Db 32505 VGEGBGHVKKYCKIENYDOSTQYTWYFGVQRLNSEKEIETTEDGVALLYKNDITKLDG 32564
QY 26 ----- 25
Db 32565 TYRCKVNDYGEDSYAELEFVGKREYDYCRRTMKIKRRDTMMLERPEFTLPLY 32624
QY 26 ----- 25
Db 32625 NKTAIVGENVRFGVTITVHEPHVTWYKSGOKIKPGDNKKYTFESDGLYQTLINSVYT 32684
QY 26 ----- 25
Db 32685 DDOAEYTVARNKYGEDSCAKLVTLHPPTDSTLRPMFKRLLANAEQOGOSVCFEIR 32744
QY 26 ----- 25
Db 32745 VSGIPPLTKWENDGQPLSLGPNIEIIEGLDYALHIRTLPEDTGYRVTATNTAGST 32804
QY 26 ----- 25
Db 32805 SCQHLQVERLRYKQEFKSEHEHRYQOKIDKTLMAELISGTEVPILYQVAKKALRE 32864
QY 26 ----- 25
Db 32865 AAVLYKPAVSUTKVGEFRLIEEKKERBLRMPYDVEPRKYKQTIIEEDQRIKQFVPM 32924
QY 26 ----- 25
Db 32925 SDMKWYKIRDOYEMPKLDRVQKRPRKIRLSHWEQFYVMPLEPRITDQYRPMKIRPKLS 32984
QY 26 ----- 25
Db 32985 QDDLEIYRPARRKTPSPDYDYFYPRRSLGDISDEBELLRIDDYLA MKKTEERLLEE 33044
QY 26 ----- 25
Db 33045 ELEIGFSAPSPSRPPELISLRYSSQAHVKEETRKDFRYSTYHPTKAEASTSYAE 33104
QY 26 ----- 25

Db 33105 LREHQAAYRQPKORQIRMAERDEELLRPVTTQHLSEYKSELDMSEKSRKRR 33164
QY 26 ----- 28
Db 33165 QREYTEITEIEEYBISKHAORESSSASRLRRRSLSPTIYELMPVSELLRSRQPA 33224
QY 29 ----- 28
Db 33225 EYEDDTERSPPTERTPRSPSPVSSERSLSRERSARFDIFGRYESMKALKTKTQTS 33284
QY 29 ----- 28
Db 33285 RRYEVLSSQPFLLDHPRIILRMRSHRVPQGNTRFILNVQSKPTAEVKYHNGVELQES 33344
QY 29 ----- 28
Db 33345 SKIHTNTSGVLTLEILDCHTDDSGTYRACVTNKGASDXYATIDYVGGDYTYASORD 33404
QY 29 ----- 28
Db 33405 EEPVRSVPPELRTAEYAVSSFKTSEMEASSVREVSQMTETRESLSYEHSAEMK 33464
QY 29 ----- 28
Db 33465 SAALFEKSLEEKSTTRIKITLLAAILTRKPRMTVYEGESARFSCDTDEPPTVTLRK 33524
QY 29 ----- 28
Db 33525 GQVLTSAHQVTTKYKSTFEISSVQASDEGNTSVVENSEGQAEFTLTIOKARYTE 33584
QY 29 ----- 28
Db 33585 KAVTSPPRVKSPEPRVKSPEAVKSPKRVKSPESHKPAVSPTETKPTTEKVOHLVPSAP 33644
QY 29 ----- 28
Db 33645 PKITQFLAASKETAKLTCVYESVLAKEVYWKDGKTKLKGHGFQFHSADGTYELK 33704
QY 29 ----- 28
Db 33705 INNLTESDGEYVCEISGBGSKTNLQFMGAFAKSIHEKYSKISEFKSDOKTTESTVT 33764
QY 29 ----- 28
Db 33765 RTEPRKAPRISSKPVIYTLQDVTSSDSVAKPAVKATGEPRTAIVTKDGKAITOGK 33824
QY 29 ----- 28
Db 33825 YKLSDEKGFLEIHKTDSDSLGYTCYKNSAGSVSSCKLTITKAIKDTBAQVSTQKT 33884
QY 29 ----- 28
Db 33885 SEITPOKAVQOEISOKALRSEIIMSEAKSOEKLALKEASKVLISEEVKSAATSL 33944
QY 29 ----- 28
Db 33945 KSIIVHEITKTSQASEEVTAETAEIKAFSTQMSINEGQRLVILKANIGATDVYKWLNGVEL 34004
QY 29 ----- 28
Db 34005 TNSSEYRYGVSSDQTLITKQASHDEGLIHCISKTKEGIYKQCYDLTLKSELSDAPAFI 34064
QY 29 ----- 28
Db 34065 SQPRQINEGQNVLTCEISGEPSELEIEMFNKLNPLISISNVSISRSRNVYSLEIRNAS 34124
QY 29 ----- 28
Db 34125 VSDGKYTTIKANFNQCSATASLWVPLVEEPSREVVLTFTSGDTSLOGSPSSQSVQMSA 34184
QY 29 ----- 28
Db 34185 SKQEASFSSFSSSASAMTEMKFASMSAQSMSMSQESFVEMSSSFNGISNNTQLESSTS 34244

Db 4557 GEDTPOGVPDGRPFANTQVYVLDLSALRPVAPGVGTGELYIAGBQLARGYLGRPALTAERF 4616
QY 17 ----- 16
Db 4617 TANPHSTPGARMYRTGDLAHMNHGHLTYDGRADHOIKLRHRIEPEETEATLTQOTGI 4676
QY 17 ----- 16
Db 4677 TQATVQLREDOQGDRLVATLVYNDSTEYDEKTYRDLATSALPDYVWPSALVTLDALPLT 4736
QY 17 ----- 16
Db 4737 PNGKIDRTALPAPAYASASTAGRAPRPREVEVLTCTFAEVLGVDTVTIDNFDLGGHLL 4796
QY 17 ----- 16
Db 4797 ATRLVSHRTALGVELSIRQLFETPTVAGLAALDASGVTRALTARPRERIPLSYAQQ 4856
QY 17 ----- 16
Db 4857 RLWELHOLEGFSATYNTVTLRLGALVDALRAAISDVARHESLRTVTEDERGAYOI 4916
QY 17 ----- 16
Db 4917 VLPVEASTPFTVYVAEEIGDRLDEAVGHCFLAQELPARTSLFVRSREHVLTLHL 4976
QY 17 ----- 16
Db 4977 HIASDANSRAPLAODLTAAVTAARVSEAPWMAPLTVQYADYALMOEILGDDTDADSLAG 5036
QY 17 ----- 16
Db 5037 RQLAYWQOLAGLPEQLDPTDRPRPAVAGYSGDRVPFTVTEHLRTELARATNTSAF 5096
QY 17 ----- 16
Db 5097 MVIOAAVAVLLTRLGAGEDIPIGTPVAGRTDADADLLGLFINTLVLRITDTSDDPTFRRL 5156
QY 17 ----- 16
Db 5157 LDRVRDIDLAAVAAHODLPFERLVALNPARTLSHPLFOVLLTFNNTHGALKDISELP 5216
QY 17 ----- 16
Db 5217 GLTVALREVQRTSSKFDLSFGFAESFDSRRPOGIEAALDSTELDRSQAIALDRFLR 5276
QY 17 ----- 16
Db 5277 VLEAVTTAPDRPGAVELMDPAERERVLVEMNGAPOTLPCTPLHELISEQARLTPDAVAV 5336
QY 17 ----- 16
Db 5337 VCDGTTLYAELDRRANOLAHLLGEGIGAEDEFAIALAKSLDAVISMLAVLKTGAALP 5396
QY 17 ----- 16
Db 5397 IDPYPAERITYMLDDAQPLLTAPIPASYSRPTSEITDVERSPWSARHAAVMIYT 5456
QY 17 ----- 16
Db 5457 SGTGRKGVVIEHHAALATYILHARANTYTMGTGTVLHSPILADLTITALTPLTAGTV 5516
QY 17 ----- 16
Db 5517 HLTSLAEAEVQPSLIKATPSHLPLLTLPETASPSHTLLIGGELAHOTDLATWRTQPGA 5576
QY 17 ----- 16
Db 5577 QIINAYGTESTVNTDHHVSEDPDGPVPIGRPFANTQVYVLDLSALRPVAPGVGTGELY 5636
QY 17 ----- 16

Db 5637 AGEOLARGYLGRPALTAERTANPHSSTPGARMYRTGDLAHMNHGHLTYDGRADHOIKL 5696
QY 17 ----- 16
Db 5697 RGHRIEPEIETTLTQOTGITTQVQLREDTPGQORLVATLVYNDSTEYDEPTLRDALAS 5756
QY 17 ----- 16
Db 5757 ALPDYMRSAVYTLDALPLTPNGKIDRTALPAPAYASASTGTGRTPTREIITCTFAEVL 5816
QY 17 ----- 16
Db 5817 GVDLVTIDNFDLGGHSLAFLVSRARATLGVLSYKOFETPTIAGLSGADFDRAGRA 5876
QY 17 ----- 16
Db 5877 RAALTARPRERIPLSYAQQRLWFLHOLEGFSATYNTVTLRLGTLDTALQSALNDLL 5936
QY 17 ----- 16
Db 5937 ARHESLRTTYEDEGEPQVTHAMEPGMLPLGVVDTEGELDAMLGAVHHAFLDTAGIP 5996
QY 17 ----- 19
Db 5997 VRATLFRISOEHVLLLIHHIADAMSRTPLGHDLAAYGARCAGDVPAMEPLPVOTAD 6056
QY 20 ----- 19
Db 6057 YALMQREVLGDEGDADAPAGRLAVWTRQLADLPEQLDPTDRPRPAVASQGDVAFSL 6116
QY 20 ----- 19
Db 6117 DADLVYRLTELARATHSSTFVYQALAVLLTRLGAGEDIPIGTPVAGRTDADATENLVGF 6176
QY 20 ----- 19
Db 6177 FVNTLVLRNDTSGNPTFRLEETRTDLAAVAAHODLPFERLVALNPARTLSHPLFOV 6236
QY 20 ----- 19
Db 6237 MLILSTAETPDASALPLGLRYGABRSRLGAKVDALEALAEVRDGEGRSTGLTGALDFR 6296
QY 20 ----- 19
Db 6297 TDLFDRSTARSLVERFVTLTLEAVADPGVRLSRVPLTGSERRSLDRGTGRLLEGIDAT 6356
QY 20 ----- 19
Db 6357 LPELFAEQALRTPGAPALVRGSTVSYAELDRTNRLRLRQGVVRPCTPVMLMERSP 6416
QY 20 ----- 19
Db 6417 AHVATLAIKAGAVVPLHDYTPILDRMRHVVAADTAALILITDRAEARAGQLGARVAV 6476
QY 20 ----- 19
Db 6477 DEFGAAPSGSEADAPGTGTGTGTSRSGYVDADAEVGLRPDOLAYVMTTSGSTGVPRGV 6536
QY 20 ----- 19
Db 6537 AVTHRGVVDLVNDHCWRPGVHERVLLHAPHADVSCYEMKVVLYSGGVVAVAPGLDPA 6596
QY 20 ----- 19
Db 6597 AITDLITADITAIHLTAGFRRVVAEEAECEPAGVREVLTGDDVVSPPAAVAVLAHHPRI 6656
QY 20 ----- 19
Db 6657 VLRHLTGPTETTLCTQHCVTAPEAKGSLPYGR 6690
QY 20 ----- 19

RESULT 6
000308

```
ID 000308 PRELIMINARY; PRT; 870 AA.
AC 000308;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE WMP2.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97313427; PubMed=9169421;
RA Pirozzi G., McConnell S.J., Uveges A.J., Carter J.M., Sparks A.B.,
RA Kay B.K., Fowlkes D.M.;
RT "Identification of novel human WW domain-containing proteins by
RT cloning of ligand targets."
RL J. Biol. Chem. 272:14611-14616(1997).
DR EMBL: U96114; AAC51325.1; -.
DR HSSP: Q13526; IPIN.
DR InterPro: IPR000008; C2.
DR InterPro: IPR000569; HECT_domain.
DR InterPro: IPR002349; WW.
DR InterPro: IPR001202; WW_Rsp5_WMP.
DR Pfam: PF00632; HECT; 1.
DR Pfam: PF00397; WW; 4.
DR PRINTS: PR00403; WMDOMAIN.
DR SMART: SM00239; C2; 1.
DR SMART: SM00119; HECTC; 1.
DR SMART: SM00456; WW; 4.
DR PROSITE: PS00237; HECT; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 4.
DR PROSITE: PS00020; WW_DOMAIN_2; 2.
SQ SEQUENCE 870 AA; 99070 MW; 63342143211D0294 CRC64;

Query Match 42.3%; Score 66; DB 4; Length 870;
Best Local Similarity 2.9%; Pred. NO. 1.3e+02;
Matches 18; Conservative 2; Mismatches 11; Indels 588; Gaps 4;

QY 1 SRXHXHS----- 7
DB 191 SRTHHSGASARTTPATGEGSPGARSRHQPYKNSGSHGLANGTVNDEPTTADPEEPSV 250
QY 8 ----- 7
DB 251 VGVTSPPAAPLSVTPNPNTSLPAPATPAEGEBPSTGTOQLPAAQAPDALPAGWQRE 310
QY 8 -----MEXRT----- 12
DB 311 LPNGRYVVDHNTKTTTWERPLPPGWEKRTDPRGRFYVVDHNTTTTQWPTAEVRYNYE 370
QY 13 ----- 12
DB 371 QMOSQNOLOGAMQHSQRLYQFMSASTDHDPLGLPPGWEKRODNGRYVYNHNTRTT 430
QY 13 -----PDINPAM--YXXRGIR----- 26
DB 431 QWEDPRTQGMIOEPALPPGWEKMYTSEGVRYFVDHNTRTTTFKDPDPGEGSTGQSPGA 490
QY 27 ----- 26
DB 491 YDRSFRWKYHQFRFLCHSNALPSHVKISVSQTLFEDSFQOIMMKPYDLRRRLYIIMG 550
QY 27 ----- 26
DB 551 EBGUDYGIGIAREWFFLLSHEVLNPMYCLFEYAGKNKYCLOINPASSINPDHLLTYFFRIGR 610
DB 611 FIAMALYHGKFIIDGFTLPPYKRLNKRPTLKDLSEIDPEFYNSIWIKENLECGLEL 670
QY 27 ----- 26
```

```
DB 671 YFIQMEILGKVTTHLKEGSESTRVTEENKEEYIMLLTDMRFTRGVEQTKAFIDGFE 730
QY 27 ----- 26
DB 731 VAPLEWLYRFDKELEMLCGMQEIDMSDMQKSTTYRHYTKNSRQIQMFWQVYKEMDNEK 790
QY 27 -----PYGRF 31
DB 791 RIRLQFVYTGCRLPVGVGF 809

RESULT 7
ID 096CZ2 PRELIMINARY; PRT; 870 AA.
AC 096CZ2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Medd-4-Like ubiquitin-protein ligase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC013645; AAH13645.1; -.
DR InterPro: IPR000569; HECT_domain.
DR InterPro: IPR002349; WW.
DR InterPro: IPR001202; WW_Rsp5_WMP.
DR Pfam: PF00632; HECT; 1.
DR Pfam: PF00397; WW; 4.
DR PRINTS: PR00403; WMDOMAIN.
DR PROSITE: PS00237; HECT; 1.
DR PROSITE: PS01159; WW_DOMAIN_1; UNKNOWN_4.
DR PROSITE: PS00020; WW_DOMAIN_2; 2.
DR KW ligase.
SQ SEQUENCE 870 AA; 98912 MW; FCCD75CBA61F2204 CRC64;

Query Match 42.3%; Score 66; DB 4; Length 870;
Best Local Similarity 2.9%; Pred. NO. 1.3e+02;
Matches 18; Conservative 2; Mismatches 11; Indels 588; Gaps 4;

QY 1 SRXHXHS----- 7
DB 191 SRTHHSGASARTTPATGEGSPGARSRHQPYKNSGSHGLANGTVNDEPTTADPEEPSV 250
QY 8 ----- 7
DB 251 VGVTSPPAAPLSVTPNPNTSLPAPATPAEGEBPSTGTOQLPAAQAPDALPAGWQRE 310
QY 8 -----MEXRT----- 12
DB 311 LPNGRYVVDHNTKTTTWERPLPPGWEKRTDPRGRFYVVDHNTTTTQWPTAEVRYNYE 370
QY 13 ----- 12
DB 371 QMOSQNOLOGAMQHSQRLYQSSASTDHDPLGLPPGWEKRODNGRYVYNHNTRTT 430
QY 13 -----PDINPAM--YXXRGIR----- 26
DB 431 QWEDPRTQGMIOEPALPPGWEKMYTSEGVRYFVDHNTRTTTFKDPDPGEGSTGQSPGA 490
QY 27 ----- 26
DB 491 YDRSFRWKYHQFRFLCHSNALPSHVKISVSQTLFEDSFQOIMMKPYDLRRRLYIIMG 550
QY 27 ----- 26
DB 551 EBGUDYGIGIAREWFFLLSHEVLNPMYCLFEYAGKNKYCLOINPASSINPDHLLTYFFRIGR 610
```

```

OY 27 ----- 26
DB 611 FIAMALYHGKFIIDGTGTLPPYKRLMKRPTLKDLESIDPEFYNSIWIKENNLEECLEL 670
OY 27 ----- 26
DB 671 YFIQDEILGKVTYTHLEKEGESINTEENKEEYIMLLTDMRTGVEQTKAPLDFGNE 730
OY 27 ----- 26
DB 731 VAPLEMLRTFDEKELMLCMQMOEIDMSDMOKSTIYRYTNSKOIQMFQVYKEMDNK 790
OY 27 -----PVGRF 31
DB 791 RIRLQFVTGTCRLPYVGF 809

RESULT 8
OY DBH0 PRELIMINARY; PRT; 870 AA.
AC Q9DBH0;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE 1300010066RIK protein.
GN 1300010066RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aitawa K., Itawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Kono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guenichon S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK004962; BAB23702.1; -
DR HSSP; Q13526; IPIN.
DR MGD; MGI:191414; 1300010066RIK.
DR InterPro; IPR000008; C2.
DR InterPro; IPR000569; HECT_domain.
DR InterPro; IPR002349; WW.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00632; HECT_1.
DR Pfam; PF00397; WW_4.
DR PRINTS; PR00403; WMDOMAIN.
DR SMART; SM00239; C2_1.
DR SMART; SM00119; HECTC_1.
DR SMART; SM00456; WW_4.
DR PROSITE; PS50237; HECT_1.
DR PROSITE; PS01159; WW_DOMAIN_1; 3.
DR PROSITE; PS50020; WW_DOMAIN_2; 2.
DR PROSITE; PS50020; WW_DOMAIN_2; 2.
SQ SEQUENCE 870 AA; 98760 MW; 72B34A1B727A7FB3 CRC64;
Query Match 42.3%; Score 66; DB 11; Length 870;

```

```

Best Local Similarity 2.9%; Pred. No. 1.3e+02;
Matches 18; Conservative 2; Mismatches 11; Indels 588; Gaps 4;
OY 1 SRXKHS----- 7
DB 191 SKTHRHSGSARVATAASEQSPGARNRHOPVKNSSSGIANGVNEEPTPASEPESSV 250
OY 8 ----- 7
DB 251 VCVTSLPAALSVSNPNTSLPAOSTPACGEASTSGTQQLPAAQADALPAGWEORE 310
OY 8 -----MEXRT----- 12
DB 311 LPNGRVYYVDHNTKTTWERPLPGWEKRDPGRGFEYVDHNTRTTQRPAAEYRANE 370
OY 13 ----- 12
DB 371 QMOSQROLOGAMQHSQRELYOSSASTDHDPLGPLPGWEKRODNGRVYVNHNTRTT 430
OY 13 -----PDINPAW---YXXGIR----- 26
DB 431 QWEDPRTQGMIOEPALPPGHEMKTISEGVRYFVDHNTRTTTRKDPAPGESGKQSPGA 490
OY 27 ----- 26
DB 491 YDRSFRWKYHQFRLCHSNALPSHKISVSQTLFEDSFQOLNMKPYDLRRRLYIMNG 550
OY 27 ----- 26
DB 551 EEGLDYGGIAREWFFLSHEVLNPMYCLFEYAGKNYCIQINPASSINDHLTYFRFTGR 610
OY 27 ----- 26
DB 611 FIAMALYHGKFIIDGTGTLPPYKRLMKRPTLKDLESIDPEFYNSIWIKENNLEECLEL 670
OY 27 ----- 26
DB 671 YFIQDEILGKVTYTHLEKEGENIRYTEENKEEYIMLLTDMRTGVEQTKAPLDFGNE 730
OY 27 ----- 26
DB 731 VAPLEMLRTFDEKELMLCMQMOEIDMSDMOKNAYRYHTKSKOIQMFQVYKEMDNK 790
OY 27 -----PVGRF 31
DB 791 RIRLQFVTGTCRLPYVGF 809

RESULT 9
OY DBH0 PRELIMINARY; PRT; 1889 AA.
AC Q94H07;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Putative phosphatidylinositol kinase.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriatoidae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Ouyang S., Mofat K.S., Hill J.N., Gansberger K.,
RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tselirin T.,
RA Riggs F., Hsiao J., Zisman V., Blunt S., Pal G., VanAken S.E.,
RA Uteback T.R., Feldblum T.V., Quackenbush J., Salzberg S.L.,
RA White O., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSJNBa0018H01 genomic sequence.";
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC087181; AAK38486.1; -
DR InterPro; IPR001263; P13Ka.

```

DR InterPro: IPR000403; PI3_PI4_kinase.
 DR Pfam: PF00613; PI3Ka; 1.
 DR Pfam: PF00454; PI3_PI4_kinase; 1.
 DR PROSITE: PS00915; PI3_4_KINASE_1; UNKNOWN_1.
 DR PROSITE: PS50290; PI3_4_KINASE_3; 1.
 DR KINASE.
 SQ SEQUENCE 1889 AA; 207787 MW; ED5828DB3A89557 CRC64;

Query Match 42.3%; Score 66; DB 10; Length 1889;
 Best Local Similarity 1.2%; Pred. No. 4.5e+02;
 Matches 22; Conservative 1; Mismatches 8; Indels 1773; Gaps 6;

QY 1 SRXHXH-----7
 DB 45 SRAHLHSLALARLLPAGAGAGSSLPAPLAFASHAFSPSPWQSFAPAPLSRLPL 104
 QY 8 -----7
 DB 105 LAAPTCALSSAALLAALDVADPASPAPLAPLSSAAAAPPTLLPADAAFPVASRL 164
 QY 8 -----7
 DB 165 LEFAGEAPPKAKGGEYAAAGEENGVRVYQKFEFEVEBELERKEVAFRLIVHVLGE 224
 QY 8 -----7
 DB 225 GGLSDNVAAYRNAAKOVNSLSEFLKIRKRDWREGAQLKTRINTKLCCQAAYVLYR 284
 QY 8 -----7
 DB 285 SVSADVDKSKKMDLQOTLAWFIETAKSCILSSWRKLICEELFGLLNVKLSCEITE 344
 QY 8 -----7
 DB 345 FGWTKRALVDFTIMRLAAYVRERNDYEEDGKEKDAVPVRLNVLRLAELCVCLKWE 404
 QY 8 -----7
 DB 405 VVDMLPLFTEHLEEGDASSPILRLRLDAISRVACLGFEKSYRDSIVLMTRSYLDKAK 464
 QY 8 -----7
 DB 465 AVGSAENNTVSEATERIETLPAGFLVATNLTSTKLSRDYHRLSLSCSDVGLAESEK 524
 QY 8 -----7
 DB 525 SGSGADLMGPLLPAVAELCSDFDPVSTVEPSLLKLFRLMFLYVLFGLAPPIQSNQTPA 584
 QY 8 -----7
 DB 585 KPVSTSLNWEISATALQAVSGPYMNSHMCVAVORIAOGTTPPLVSSVYKMLEDELEIN 644
 QY 8 -----7
 DB 645 ALHNPGRGNSNEKAAGVGTALSAALGSRYEVAAMTISGVKATYLLAVALFELIRFS 704
 QY 8 -----11
 DB 705 CNGIILSTATSTLNKSASFCEYELLITNLPLPAVSQCLTAVVHRAFTVLSMEDRID 764
 QY 12 -----11
 DB 765 IGRGADIRSVISVHACFLIKSMQDENVRDYSVKLQLKEKPPQVLAHSSCVDLILI 824
 QY 12 -----11
 DB 825 SVHNEITSGVSDPAAVAVTSLYOKIAREWITSALSTAPCTTQGLIDENFCKPSGAQRS 884
 QY 12 -----16
 DB 885 QHTADVLSLSEIRICSGKNMGITFANVPVAMDSAAAASAKAKAPITTEVLSTAVV 944
 QY 17 -----16

DB 945 TATVKNHAGEIAGMRRLPSSMGAMTGMSPLEGASAPNOSFDEVLSRFVRLLODFVY 1004
 QY 17 -----16
 DB 1005 TAEKNQIDNTVFRETCSQSTALLIDHMSDSRANLDGFSQILRLCWPATYICTPDAMET 1064
 QY 17 -----16
 DB 1065 GIFWTWLVSAAPSLGPLYLAELVDAMLTIDTKGGLFASDMNMGCPDAKLRLPHLISGEP 1124
 QY 17 -----16
 DB 1125 EAPPEKDPVEATIAHRLMGLFFIDREFVVRHDSIEQLLIGMLQGTMSKATFESHHPAA 1184
 QY 17 -----22
 DB 1185 TGTFETAMLGKFCSCQSQSNLQKCMGLQLLEDRVYRALGWFAVAPEWESQNKSPA 1244
 QY 23 -----22
 DB 1245 QREAOVSITFVHCLONERPSGASDAPKSGREGEPMLDONHPVGSVDNYTTRERERK 1304
 QY 23 -----22
 DB 1305 QLLMLSQNEADRLVYMAQPIINTKDAATFRGKISSDKWIDHAKTAFAPDPRILASIMRF 1364
 QY 23 -----22
 DB 1365 PTNSALSSBITQVOTHILELNTIPEALPFTTAKVADENSLLQOLPHMAPCSVTQALE 1424
 QY 23 -----22
 DB 1425 FLPPYKGHPRVMAVLYRLVETTPETVTFMPOLVOSLRTDDKLVEGYLLGAARRSNI 1484
 QY 23 -----22
 DB 1485 FAHILMLQGECPDEPGKEAAPKATAFHSLPAVREKIYDFTPEARMFEREFEEF 1544
 QY 23 -----26
 DB 1545 DKVTSISGVLPFLPKERRAGIKRELEKITYPGDDLPTATNKFVRISQVDSGIPLOSA 1604
 QY 27 -----26
 DB 1605 AKVPIITFENVVDKDPNDVYKQACIFKVGDDCRQVIALQVIALRLDIFQAVGLNLYL 1664
 QY 27 -----29
 DB 1665 FEYGVLPTEPERGILLEVVPNTRSRNOMGETTIDGILLETFOQDYGVPVGSPEAREMFMI 1724
 QY 30 -----29
 DB 1725 SSAGYAVALSLLOPKDRKNGNLDPDSHGRVLIHIDGFTLEISPGNGGFESAHKLSHEM 1784
 QY 30 -----29
 DB 1785 TQLDPSGTMRSDWNGPLRLCVKGYLAGRRHNGIITTVLWAMDGILPCSRSEPIANL 1844
 QY 30 --RF 31
 DB 1845 RKRK 1848

RESULT 10
 ID 042181 PRELIMINARY; PRT; 4578 AA.
 AC 042181:
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE PKD1 protein.
 GN PKD1.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_Taxid=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97449170; PubMed-9285785;
 RA Sandford R., Sgotto B., Aparicio S., Brenner S., Vaudin M., Wilson R.,
 RA Chisnoe S., Pepin K., Bateman A., Chothia C., Hughes J., Harris P.;
 RT *Comparative analysis of the polycystic kidney disease 1 (PKD1) gene
 RT reveals an integral membrane glycoprotein with multiple evolutionary
 RT conserved domains.";
 RL Hum. Mol. Genet. 6:1483-1489(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Vaudin M.;
 RT "The sequence of Fugu rubripes 259C6.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Washu;;
 RT "A Fugu rubripes Genome Sequencing Project.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF013614; AAB86683.1; -;
 DR HSSP: P98161; 1B4R.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR001024; Lipoxigenase_LH2.
 DR InterPro: IPR000483; LRR_Citerm.
 DR InterPro: IPR002859; PKD/REJ-like.
 DR InterPro: IPR000434; PKD_1.
 DR InterPro: IPR000203; PKD_cys_rich.
 DR InterPro: IPR000601; PKD_domain.
 DR InterPro: IPR002889; WSC.
 DR InterPro: IPR000130; Zn_MTPeptidase.
 DR Pfam: PF018925; Gps; 1.
 DR Pfam: PF00059; Lectin_C; 1.
 DR Pfam: PF01463; LRRCT; 1.
 DR Pfam: PF00801; PKD; 15.
 DR Pfam: PF01477; PLAT; 1.
 DR Pfam: PF02010; REJ; 1.
 DR Pfam: PF01822; WSC; 1.
 DR PRINTS: PR00500; POLYCYSTIN1.
 DR SMART: SM00034; CLECT; 1.
 DR SMART: SM00303; GPs; 1.
 DR SMART: SM00308; LH2; 1.
 DR SMART: SM00082; LRRCT; 1.
 DR SMART: SM00089; PKD; 13.
 DR SMART: SM00321; WSC; 1.
 DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
 DR PROSITE: PS50093; PKD; 9.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
 DR SEQUENCE 4578 AA; 504585 MW; 62AA07B61C4CFC7A CRC64;
 SO

Query Match 42.3%; Score 66; DB 13; Length 4578;
 Best Local Similarity 0.68; Pred. NO. 1.9e+03; 8; Indels 2950; Gaps 4;
 Matches 17; Conservative 6; Mismatches

QY 1 SRXHXSM----- 9
 Db 466 ARKTHSERGVWLGIDVNSAGKLHWNGSEAGEGGLAPSSVARGNLCVSLDQRSQ 525
 QY 10 -----XRPD----- 14
 Db 526 TSSHCNARAYVCQYNQVYRVDGVSFAMGLAVFPSHLLQATAMITPPPPHGGVEV 585
 QY 15 ----- 14

Db 586 LLPALSFVRAGRLSIEFVYRELSSQTHILFOIYRECRYPGCGGACSPMALCVPSSES 645
 QY 15 ----- 14
 Db 646 RSADPPSCPILQMCPPQHRCPLPSSPCQSSCLNCQGHLPSEAMPILYTIQYEVLF 705
 QY 15 ----- 14
 Db 706 LTPGAHVLLQDQLEDLVSRGDVIALQHDGAPSLIHGQSPHSLMRQPFALNISEW 765
 QY 15 ----- 14
 Db 766 FGINTPVPPDNLADGADPLPDELKALVEGTRGSMENAVCPILRYLYGQSETRLR 825
 QY 15 ----- 14
 Db 826 GAELSAGLPQPGILNVLVTSKQSDLVASCSIQVLPPLDLTVLHSPQNGTLFLESNT 885
 QY 15 ----- 14
 Db 886 RLLLRVQRYETKAFYRTSNSSATFHPYCPQEFSSRLCHSTSRGSGAEVAERSMYAVLD 945
 QY 15 ----- 14
 Db 946 LNLKIKHTGPVQVELMAHNNVTASLTVLQLEESLGLVQPHPARRYLMEVSYTA 1005
 QY 15 ----- 14
 Db 1006 SYLEGSNPTFKMYVDKPFYTYNTVLNVIYQHADVYKLVATANNHYSTLLEDENVTVDR 1065
 QY 15 ----- 14
 Db 1066 LQPMNLTVKGVDPVPOGSTQTLTTSVLIDMAVPATVMSFGDGGYEQEPHKPPYSSL 1125
 QY 15 ----- 14
 Db 1126 TCPESPVOVLLSNNTYIYSGPIYTAVSVSNRENISRSINMSYSILTRVDIQTEPS 1185
 QY 15 ----- 14
 Db 1186 LLLAGKTADFEMHPLPSFYGIHEWNPFGSALLQGRVATFPKSGPFNVCSVNTTIS 1245
 QY 15 ----- 14
 Db 1246 STAVCTEMFAVEIDSLAESSTPELHSATYRAHLSGNNITWTSMGDKSYTISEP 1305
 QY 15 ----- 14
 Db 1306 TVSHKAYADGRTVNTLMANVSGWTIIPVEVFQYVAGIEPSGVEHALVNFRAV 1365
 QY 15 ----- 14
 Db 1366 GNASSIYEMSPGDSPNETHHNGPISHTYKNGTHPLSLITSGSKANFTQVCVOP 1425
 QY 15 ----- 14
 Db 1426 VLFKISVTAEKSHFAVGEKIQQAAREPENTTYEMDGEEDLVLPHPANVYKTYNPP 1485
 QY 15 ----- 14
 Db 1486 GCYVTVAAASNNVSNVTGSIIEVLTPYGAVYFQNGHTHNNLLGVPYFSAFSSASNV 1545
 QY 15 ----- 14
 Db 1546 SYLMNFGHNLKGNILATFNTPGHNNITLTAANRGNKHTTVPAVALASVGLTINS 1605
 QY 15 ----- 14
 Db 1606 LNVVPLNTSVHEBAHMDGDSVRFSGIILCDCTPIFRNNTMYTTRSVGTENIITYAEND 1665
 QY 15 ----- 14
 Db 1666 VGNAGASTILFVQRELEGLIIVETAEGGGGVQELDMCYFETNRVVRFHAGFKEGTNLT 1725


```

QY 15 -----INPA----- 18
Db 1726 FTWVILKNEPDNISIFNLGTAKVNPSPGCDIFLOANILQGVYVNTITYLEPARN 1785
QY 19 ----- 18
Db 1786 VOLOISNPVAVNALINNTVLMNBSNLQYRMFVNGDLDQMSKMSHTTISAGOKOVY 1845
QY 19 ----- 18
Db 1846 KVFNEVSSEVSEIYSVOEVLGKITSDATEQNYPTVDVSVCLQGBLSTGTNTVSWL 1905
QY 19 ----- 18
Db 1906 IDGSKMOKKACVTPPKPTVAVTLNATNDVSGKVSREFFVQERIFSLKASKKIAAI 1965
QY 19 ----- 18
Db 1966 NEKVEFSISLVAGTNNVLSISISDADVVLQPNQTYVHTSRVDYVAVNLTAVNEVSKR 2025
QY 19 ----- 18
Db 2026 RNLHIEVMEPYRGLSIDSCAIAVGEKVPVAVNQTGKPVHPLMTFDLHHHKEHMGK 2085
QY 19 ----- 18
Db 2086 EVSYTAEEGLITTYGAINALHAQNTTROMLAQNLMDAVLYALPDQTEVNNKVTANVL 2145
QY 19 ----- 18
Db 2146 VTPKSNMDCIMIEGDYSAPRHTNTNTVGHETTHPGHYRQVQNCNLSVWSAHAETNVS 2205
QY 19 ----- 18
Db 2206 VLECEBEVQVQAPRLAIMRSQPTLVEASVDLKGCLRGAQYLMQLLSAPSCNDPHEA 2265
QY 19 ----- 18
Db 2266 SGRVNGATRSFPVLPDAEVDVQRLQLSLPKMLAAGNTLVSLSEYENPLKKAQLOL 2325
QY 19 ----- 18
Db 2326 GVMARLPIIEGTYVWMSRTQDLOLSAQOSYDPNNDPNDOSLIHHECOSTSKGPEH 2385
QY 19 ----- 18
Db 2386 CSTLSFGLSGKGPVYLGISGELEAGIEYTFKLISGEGMPPESTQTVLVQSGHIPVYL 2445
QY 19 ----- 18
Db 2446 ECVSCRAQSLYEVSQNSYVTLRGTCSCNCGCFHGRMSAVTLQNDTLVLDSSTTTGSGM 2505
QY 19 ----- 18
Db 2506 NLVLRQGVHLHGSYNFTLAVTDDSLDGECAASITLHHNMPDGGGCHLRGGEFTGOEHG 2565
QY 19 ----- 18
Db 2566 DKEDEVRRIETLLDRVQFSGSGYSDLGVSFETPLLYSLVNRCHDYCEEFYVRSSEPH 2625
QY 19 ----- 18
Db 2626 SAFPPEFSSAQHRVSVFTVEDHOGAATAALKTKIVLPDPPPEYSSLPMLSLIDS 2685
QY 19 ----- 18
Db 2686 KLLKLDQGFQFVRELISLVLTVLNEYEQTRSVRSVRSERGRVNNITRALTAID 2745
QY 19 ----- 18
Db 2746 LTTVNDIQOTSAALOCTAVSREFICECONSTLNKLESMLTLLQDTROGIATVPTREIAD 2805

```

```

QY 19 ----- 18
Db 2806 NIINIMGDLIHQVSQSTSQSYNPPYVDSPSPFIOKMGGLDPPSPSLPHPLRYAARA 2865
QY 19 ----- 18
Db 2866 YSLSVLMLILMARVILNEEPLVKGAEIAATGKLADPQSLCYHGNNSPGFTLHETLK 2925
QY 19 ----- 18
Db 2926 ANILFLSSRLALLFSFHQVESNPPEFVYVANYTVSTEVASMEFTENGTHIPISGLD 2985
QY 19 ----- 18
Db 2986 SLATVAVNNGSNGAETGPEGAGTGVPITAGAVNISCDSYIVRVSMTNRRAGLEVOL 3045
QY 19 ----- 18
Db 3046 NFTSLQDVGLQDKREAEELITAYLHSHKPNFNCOTDKRRTLTGMRGHDLDHRYTF 3105
QY 19 ----- 18
Db 3106 FLSPELYDTTLDYFTIVSTACSGSGAGVRLVGVFASLQYFSESEKQMTDQWPLA 3165
QY 19 ----- 18
Db 3166 ETNASRVCRTRLTLFAAGLFPANASFTYPERSGTSLVLYLVYGLMSYVAAAI 3225
QY 19 ----- 18
Db 3226 LKHLQDLDRRAVAVPLCGCDGLFKYEIQKTGWSGAGTFAHVGISLYGRESRGHRL 3285
QY 19 ----- 25
Db 3286 DSRGTFARNAALDFHATDTSLGWNWKARIMHDKNGLSPAMLYQVLYKDLQTSSTYFL 3345
QY 26 ----- 25
Db 3346 VEEMLSDNEKTDGHEVEVEASEAMLLQPLRLRYELQALCESHMLSLFORPPSP 3405
QY 26 ----- 31
Db 3406 FTRLOATCCALLDILLANTLMTYSIVVDVRSYSPRPSRF 3446

```

RESULT 11

010466 PRELIMINARY; PRT: 26926 AA.

AC 010466: 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Ttlh, heart isoform N2-B (EC 2.7.1.-) (Connectin).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_TaxID=9606;

ON [1]

RP SEQUENCE FROM N.A.

RC TISSUE=HEART;

RX MEDLINE=96026330; PubMed=7569978;

RA Labelt S., Kolmer B.;

RT "Tlins: giant proteins in charge of muscle ultrastructure and elasticity.";

RL Science 270:293-296(1995).

RP [2]

RP SEQUENCE OF 22277-25376 FROM N.A.

RX MEDLINE=92258380; PubMed=1582406;

RA Labelt S., Gautel M., Lakey A., Trinick J.;

RT "Towards a molecular understanding of titin.";

RL EMBO J. 11:1711-1716(1992).

RP [3]

RP SEQUENCE OF 1976-2014 FROM N.A.

RA Label: S;
 RT Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP CHARACTERIZATION.
 RA MEDLINE-9531314; PubMed-7607248;
 RT Gautel M., Castiglione-Morelli M.A., Pfuhl M., Pastore A.;
 RT "A calmodulin-binding sequence in the C-terminus of human cardiac
 RT titin kinase."
 RL Eur. J. Biochem. 230:752-759(1995).
 CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
 CC ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
 CC SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
 CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
 CC PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
 CC DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY
 CC ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM
 CC N2-B.
 CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
 CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
 CC KINASES.
 CC -1- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
 CC IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
 CC DOMAINS.
 CC EMBL: X64698; CAA45939.1; -
 CC EMBL: X63270; CAA58243.1; -
 CC EMBL: X64697; CAA45938.1; -
 CC EMBL: X90568; CAA62188.1; -
 CC EMBL: X64699; CAA45940.1; -
 CC HSP: P56276; ITIK
 CC InterPro: IPR000282; Cytok_receptor_2.
 CC InterPro: IPR000719; Euk_kinase.
 CC InterPro: IPR000577; FGGY_kin.
 CC InterPro: IPR003962; FNIII_repeat.
 CC InterPro: IPR003961; FN_III.
 CC InterPro: IPR001097; HLH_basic.
 CC InterPro: IPR003598; Ig_C2.
 CC InterPro: IPR003600; Ig_Like.
 CC InterPro: IPR003006; Ig_MHC.
 CC InterPro: IPR000129; Peptidase_S24.
 CC InterPro: IPR002016; Peroxidase.
 CC InterPro: IPR002290; Ser_thr_kinase.
 CC InterPro: IPR001245; Tyr_kinase.
 CC Pfam: PF00041; fn3; 132.
 CC Pfam: PF00047; Ig; 91.
 CC Pfam: PF00069; kinase; 1.
 CC PRINTS: PR00014; FMYPEIII.
 CC PRINTS: PR00726; LEXASERPTASE.
 CC PRODOM: PD000001; Euk_kinase; 1.
 CC SMART: SM00060; FN3; 127.
 CC SMART: SM00408; IgC2; 23.
 CC SMART: SM00410; Ig_Like; 79.
 CC SMART: SM00220; S_TKC; 1.
 CC PROSITE: PS00933; FGGY_KINASES_1; UNKNOWN_1.
 CC PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 CC PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 CC PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
 CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
 KW Serine/threonine-protein kinase; Alternative splicing; Repeat;
 KW Immunoglobulin domain; Phosphorylation.
 FT DOMAIN 1370 1389
 FT 4 X 5 AA TANDEM REPEATS OF R-N-S-P-A.
 FT 4429 4614
 FT CATALYTIC.
 FT DOMAIN 24731 25070
 FT 25030 25056
 FT CALMODULIN-BINDING.
 FT MOD_RES 1372 1372
 FT PHOSPHORYLATION (BY PDK) (POTENTIAL).
 FT MOD_RES 1372 1372
 FT PHOSPHORYLATION (BY PDK) (POTENTIAL).
 FT MOD_RES 1382 1382
 FT PHOSPHORYLATION (BY PDK) (POTENTIAL).
 FT MOD_RES 1387 1387
 FT PHOSPHORYLATION (BY PDK) (POTENTIAL).
 FT MOD_RES 26171 26171
 FT PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 26178 26178
 FT PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 26184 26184
 FT PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 26190 26190
 FT PHOSPHORYLATION (POTENTIAL).

FT CONFLICT 22277 22277 T -> P (IN REF. 2).
 FT CONFLICT 22449 22449 E -> G (IN REF. 2).
 FT CONFLICT 22454 22454 T -> Q (IN REF. 2).
 FT CONFLICT 23324 23324 S -> L (IN REF. 2).
 SQ SEQUENCE 26926 AA; 2993428 MW; D5EECD3254DF523 CRC64;
 Query Match 42.3%; Score 66; DB 4; Length 26926;
 Best Local Similarity 0.1%; Pred. No. 3.4e+04;
 Matches 22; Conservative 3; Mismatches 6; Indels 24391; Gaps 7;
 1 SKXKX-----
 Db 2459 NRTASDEGPKYKLYGAVETNCNLSVEKIKIRGLDICTETQNVVFEVLSGIDVL 2518
 : 1
 QY 6
 Db 2519 WNFKEIKPSSKYIEAHGRIYKLYVNMKKDEGKTFYAGENMTSGKLTVAGAIK 2578
 QY 6
 Db 2579 PLTDQVAESQRAVEPECEVANDPSKGEWLIDGKHLPLTNIRSDGHRRLIAATKLD 2638
 QY 6
 Db 2639 DIGEYTVAVANSKISAKLKVEAVKIKTKLMLTETODAVFYELIHPNVKGVWIKNG 2698
 QY 6
 Db 2699 VVLESNEKALSVKGTISLRKIKCAIYDESYYGRGLGASARLHVEYKIIKKPDV 2758
 QY 6
 Db 2759 TALENATVAEVSVDYVPVKFHKVSEIKPSDKHRLVSEKRYKMLQNISSDAGEY 2818
 QY 6
 Db 2819 TAVGOLCKAKLFEYELIHTTKNIEVEPKTASFECEYSHFVPSMILKNVEIEMS 2878
 QY 6
 Db 2879 EKFKIVQGLHLQILIMNTSDSAEYTFVCGNDQVSAITLVTPIMTSMKDIAEKD 2938
 QY 6
 Db 2939 TTFEVTVYEGISYKWLKNVEIKSTDCQRTKRLTHSLIRNVHFGDAADYTVAGK 2998
 QY 6
 Db 2999 ATSTATLYVEARHIERKHIKIDIVLEKRRAMECEVSEPDITVQMMKDDQELITRIK 3058
 QY 6
 Db 3059 IOKEKYVRHLIPSRMSDAGKTYVAGVNSTAKLFEGRDVRIRSIKVEYVLEKORA 3118
 QY 6
 Db 3119 VVEEVNEDVDAMWYKDGIEINFOVERHKYVERRIHRMFISETRSDAGETTVAGR 3178
 QY 6
 Db 3179 NRSVTLVYNAPEPPQVLELPVTVQSGKPARFCAMISGRFPKISWYKEQLLSTGFK 3238
 QY 6
 Db 3239 CKFLHDQETVTLILTEAPEDAVYTCENKNDYGVATYSASISVEVPEVSPDEMPYP 3298
 QY 6
 Db 3299 PALITPLDQTVSEGPAPFQCRVSGTDLKVSWSKDKIKRSRFRMTQFEDTYQLLEIA 3358
 QY 6
 Db 3359 EAVPEDEGTYTFVANNVAGVSTANLSLEAPESILHERIEQEIEMKEKESFSSFLSABE 3418

QY 6 ----- 5
Db 3419 EGLSHAELOJSKINETLELSESPVPTKPDSEKEGTGPITKREVNADISMGOVATLSV 3478
QY 6 ----- 5
Db 3479 TVIGTPKTIOWFENGVLTPSADYKVFVGDHSHLILFTKLEDEGEYTCMASNDYKKT 3538
QY 6 ----- 5
Db 3539 ICSAYLKINSKGKHDTETESAVAKSLKIGRCPHFLKELKPICAOGLAIFETV 3598
QY 6 ----- 5
Db 3599 VGEPAFTVTFKENCOLCTSVYTYIHNPNSGTFIVNDPQREDSGLYICKAENMLGESP 3658
QY 6 ----- 5
Db 3659 CAHELVLLEDTMTDTPCKAKSTPEAPEDPQTPKGAVALDSEQETATVKTILK 3718
QY 6 ----- 5
Db 3719 AALITEENQOLSYEHIAKANELSSQLPGAQELQSLQEDKLPPESTREFLCINGSIHQ 3778
QY 6 ----- 5
Db 3779 PLKEPSPNLOLOIVOSQKTESKGIIMPEPETQAVLSDTEKIFPSAMSTEQINSITVER 3838
QY 6 ----- 5
Db 3839 LKTLAPPEGNYPQSSIEPPMHSTYLSVAEVLSTKEKTYSDTNREORYTLOKQOASAL 3898
QY 6 ----- 5
Db 3899 ILSQSLAEHVESLOSQDVMIQVNTPEPLVPSHESCTEGKILIESANPLENAGDSAVR 3958
QY 6 ----- 5
Db 3959 IEBGKSLRFLALKEKOVLLKEHSDNVMPPOITIESKREPAIKKVOEVOGRDILSRE 4018
QY 6 ----- 5
Db 4019 SLISGIEQRLMKIQICRALQAAVASEQPLFSEMLNIEKVEVEAVNITQEPHIMC 4078
QY 6 ----- 5
Db 4079 MYLVTSKSVTEEVYIIEDVDPOMANLKMELDALCAIYEEDILTAEGPRIQOAKT 4138
QY 6 ----- 5
Db 4139 SLQEMDSPSGQKVEPIPEPEVESKYLSTEVSYFNQSRVXYLDATPVTKVASAVV 4198
QY 6 ----- 5
Db 4199 SDEKODESLKPESEKESSESSESTEVEATVKIQEAGGLIKEDGPMIHTPLVDYSEED 4258
QY 6 ----- 5
Db 4259 IYVLTSTNAKEVVMYFENKLVPSDEKFCLODQNTYTLVIDKNTEDHGEYCEALN 4318
QY 6 ----- 5
Db 4319 DSGKATSAKLIVVKRAAVYIKRKIEPLEVALGHLAKFTCEIOSAPNVRFQKAGREY 4378
QY 6 ----- 5
Db 4379 ESDKCSIRSSKYISLEILRTQVDCGEYTCRASNEYSVCTAULTVTPGEEKVRL 4438
QY 6 ----- 5
Db 4439 LPERKPEKEVEVVLKSVLAKRPEEBEPKVEPKLEKVKPAVPEPPPKVEVEVEVPTV 4498

QY 6 ----- 5
Db 4499 KREPKIPEPTKVEIKPAIPALPAPPEPKPEAEVKTIKPPVEPEPTPIAADVTPVVGK 4558
QY 6 ----- 5
Db 4559 KAEKAPKEEAAKPKPGPIKGVPKTPSPLEABRRKLRPSSGGEKPPDEAPFTYQKAVPL 4618
QY 6 ----- 5
Db 4619 KFYKEIKDIIIESEFVGSSAIFECVLSESTAITTWMKGSNIRSPKHFADGDKRL 4678
QY 6 ----- 5
Db 4679 HIIDVQSDAGHYTCVLRIGNKEKSTAKLVVEELPVPRVKTLEBEVTVKGOPLYISCE 4738
QY 6 ----- 5
Db 4739 LNKEDVVRKDKIYVEKGRIVPGVIGLMBALITINDADDTAGTYTVYENANNLECS 4798
QY 6 ----- 5
Db 4799 SCVYVEVIRDMVLRIRDOHYKPKGTALFACDIKADTPNIKWFQYDEIPAEPNDKTEI 4858
QY 6 ----- 5
Db 4859 LRDGNHLYLKINAMPEDIAEYAVEIEGKRYPAKITLGERVELLKPLEDVTIYKESAS 4918
QY 6 ----- 5
Db 4919 FPAISEADIPQWMLKGBLLRSPTECEIKAGKRFYTLHKVKLDQAGEVLYQALNAIT 4978
QY 6 ----- 5
Db 4979 TAILVKEIELDPALVDVYPERQARFECVLTREANVIMSKGPDIIKSSDKFDIAD 5038
QY 6 ----- 5
Db 5039 GKHLIVINDSQFDEGVYTAVEEGKTSARLVYIGRLKFMSPLEQTVAGEGTATFVC 5098
QY 6 ----- 5
Db 5099 ELSHEKMHVWFKNDAKLHTSRTVLISSEKTHLEMEVTLIDISQIKAOVKELSTAQ 5158
QY 6 ----- 5
Db 5159 LKVLADPYFTVKLHDKTAVEKDEITLKCEYSKOVYKWFMDGEBEIVSPKYSIKADGLR 5218
QY 6 ----- 5
Db 5219 RLKIKKADLMDGEBYVDCGDTKRTKANVTEARLIEVEKPLYGVEYVGETAFIEILS 5278
QY 6 ----- 5
Db 5279 EPDVHGQWKLKGQPLTASPDCIILEDGKKHILLHNCOLGHTGEVSPQANAKSAAMLKV 5338
QY 6 ----- 5
Db 5339 KELPLIFITPLSDVVFEBKDEAKFECBEYSREPKTFMMLKGTQELTGDPRFELINDGTHS 5398
QY 6 ----- 5
Db 5399 MVISAFEDBAKYWFEADKHTSGKLIINGIRLKLPLKLDVYAKKESAVFVELSHD 5458
QY 6 ----- 5
Db 5459 NIRVWFKNQDRLHTRSVSMQDBGKTHSTFKDLSTDDSQIVVEAMGSSEAKLTYLE 5518
QY 6 ----- 5
Db 5519 GDPYFTGLDQYTGVEKDEVILQCEISKADAPVWFKQKEIKPSKNAVITKDQKKMLI 5578
QY 6 ----- 5

Db 5579 LKRLKSDIGQYICDGTGDKTSKGLDIEDREIKLYRPLHSEVMEIETARFETIEISDDI 5638
QY 6 ----- 5
Db 5639 HANWKLGEALLQTPDCEIKEEGKIHSLVLHNCRLDQGTGVFOAMNVKSSAHLRVKPRV 5698
QY 6 ----- 5
Db 5699 IGLRLPLKDVYTAGETATDCELSYEDIDIVEMTLGKGLPEPSDKVYPRSEKGVHLLTLR 5758
QY 6 ----- 5
Db 5759 DVKLEDAGEVQLTAKDKFTHANLFVKPEYFTKPLEDQIVEGATAVLECEVSRENAKY 5818
QY 6 ----- 5
Db 5819 KWFKNQTEILKSKKKEIIVADGRVRKLVIHDTPEIDITYCDAKFKTSCNLNVPPHVE 5878
QY 6 ----- 5
Db 5879 FLRLPLDQVREKEMARPECELSRENAKVKWFKDGAETIKGKKYDIISGAVRLVYNKC 5938
QY 6 ----- 5
Db 5939 LLDDEAEYSCVETARTSGMLTVLEEAVFTKNLANIEVSETDIKLVCEVSKPGEVW 5998
QY 6 ----- 5
Db 5999 YKGBEIIETGRYEILTEGRRRLIVIQNAHLEDAGNANCLRSPSRDQKVKVHELAAFT 6058
QY 6 ----- 5
Db 6059 SKPQNLIELBEKAFVCSISKESFPVQWKRDKTLSEGDKYDIADGKKRVLVVKDRTL 6118
QY 6 ----- 5
Db 6119 QDMGYVVMGANAHAHLIVIEKRLIYVPLKDTVRKEQOEVENCEVTEGAKAKWRN 6178
QY 6 ----- 5
Db 6179 EBAIFDSKTYIILOKDLVYTRIRDAHLDDQANVNSLTNHRGENKSAANLYEEDDLR 6238
QY 6 ----- 5
Db 6239 IVEPLKDIETMEKKSVPFWCAVNLNVTLKWTNGEEVFPDNRVSYRDKYKMLTJKDC 6298
QY 6 ----- 5
Db 6299 GEPDEGYITVAGODKSVALLIIEAPTEVEHLEDQTVTEFDDAVFSQOLSREKANVKW 6358
QY 6 ----- 5
Db 6359 YRNGREIEGKKYFEKDGSIHRLIKDCRLDDECEYACGVEDKSRARLFVEEIPVEII 6418
QY 6 ----- 5
Db 6419 RPPQDILEAPGADVFLAELNKDKVEVQWLRNNNVVVGDRHOMSEKIHRLQICDKP 6478
QY 6 ----- 5
Db 6479 RDQGEYRIANDKEARAKLELAAPKIKTADODLVVDYGRPLTMVVPYDAYPKAEAEWFK 6538
QY 6 ----- 5
Db 6539 ENEPISTKITDTAEQTSFRILEAKKGDKGRKIVLQNHKAGAFINLKYIDVPGPYRN 6598
QY 6 ----- 5
Db 6599 LEVETFDGEVSLAWEPLTDGSKIIIGYVVERDIKRRTWLATDRAESCEFTVTGLQK 6658
QY 6 ----- 5

Db 6659 GGEYELFRVSANRRVGTGEPVETDNPVEARSKYDVPGLNVTITDVNRGVSJTWPEPE 6718
QY 6 ----- 5
Db 6719 YDGAELITNYVELDKTISIRMDTAMTVRAEDLSATYVDVVEGQEYSPRVAQRIGVKG 6778
QY 6 ----- 5
Db 6779 PSAATPPVKVADPIERPSPVNLJTSDDQOSSVLKWPPLKDGSPILGYIIEECCEGK 6838
QY 6 ----- 5
Db 6839 DMWIRCNKMLVBELTYKYTGLEKGNKYLRYVAENKAGVSDPSILLGLTADDAFVEPTM 6898
QY 6 ----- 5
Db 6899 DLSAFKDLGVIVPNPITILVPSIGYPRPTATWCGRGVLETDGRVKKMTLSAYAEIVIS 6958
QY 6 ----- 5
Db 6959 PSERSDKGYTLKLENRVKTSIGEIDVNIAPSAPELKFQDITKDSVHLTWEPDDG 7018
QY 6 ----- 5
Db 7019 GSPLTGYVEKREVSRTKWTVMDEFTDLETPDLVQGEYLFKVCARNKCGEPAYV 7078
QY 6 ----- 5
Db 7079 DEPVNKNSTPATYVDPDENKWMRDTANSIFLWDPFKNDGSGRIKGIYVERCPRGSDKW 7138
QY 6 ----- 5
Db 7139 ACSEPAETKMEVTLGEBGKWAYRVKTLNRQASKPSRPEIQAVDTOEAPEIFLDVK 7198
QY 6 ----- 5
Db 7199 LLAGLTAKAGTIELPATYTGKPEPKITTKADMILKQDKRTIENVRKSTVYIVDSKR 7258
QY 6 ----- 5
Db 7259 SDIGTYIEAVNVCGRATAVEVENVLDKPPAPADITDVTNESCLTMNPPRDDGSKI 7318
QY 6 ----- 5
Db 7319 TNYVERRATDSEWHKLSITVKDITNFKATKLIIPNKEYIFRVAENMYGAGEPVQASPT 7378
QY 6 ----- 5
Db 7379 AKYQDPGPPRTLRLEPDDITKDAVTLTWCEPDDGSPITGYWVERLDPTDKWRCNM 7438
QY 6 ----- 5
Db 7439 PVKDTTYRVKGLNKKYRFRVLAENLAGPKPSKSTEBILIKDPLDPWPGRKPTVRDV 7498
QY 6 ----- 5
Db 7499 GKTSVRLNMTKPEHGDGANIESYIEMLTGTDEWVRVAEGVPTQHLGLMGOEYSF 7558
QY 6 ----- 5
Db 7559 RYRAVNKAGSESPSPVLCREKLYPPSPRWLEVINITNTADLKWTVPEKDGSPIT 7618
QY 6 ----- 5
Db 7619 TNYIVERDVRKGMQTVDTVADTKCTVPLTEGSLYFRVAENALGQSDYTEIEDSV 7678
QY 6 ----- 5
Db 7679 LAKDTFTTGPYPYALAVDVTKRHVDLKWPPKNDGGRPIQRYVIEKKERLQTRWYKAG 7738
QY 6 ----- 5
Db 7739 TAGPDCNFRYTDVIEGTEVQFVRAENAGVGHSPTEILSIEDPTSPSPPLDLHVTD 7798

QY 6 ----- 5
Db 7799 AGRHIALAMKPPKNGSPLIGYHMCPCVTEKMMRVNSRPIDKLFKEVGBVVDKE 7858
QY 6 ----- 5
Db 7859 YVLRVRAVNAIGVSEPSSEISENVAKDPDOCKPTIDLETHDIIVIEGKLSIPVFPRAVPV 7918
QY 6 ----- 5
Db 7919 PTVSMHDKREKASDRLTMKNHISAHLEVPKSVRADAGITITLENKLSATASINVK 7978
QY 6 ----- 5
Db 7979 VIGLPGCKDIKASDITSSCKLTWEPPEPDGPTPLHYLERREAGRTYIIVMSGENK 8038
QY 6 ----- 5
Db 8039 LSMTVKDLIPNGEYFFRVKAVNKVGGEYIELKNPYAQDPKOPDPDPVUVEVHNPTAA 8098
QY 6 ----- 5
Db 8099 MTTWKPEPLDYGSKIMGYIIEKIANGEHRMKRCHNEHLPILTYTAKGLEGEKEYOFRV 8158
QY 6 ----- 5
Db 8159 AENNAIGSEPSRATPTPKAVDPIDAPRVILRTSLEVKGDEIALDASISGSPYPTITWIK 8218
QY 6 ----- 5
Db 8219 DENVIPEEIKKRAAPLVRRRRKGEVOEEEPVLPLTORLSIDNSKKGESOLVRDRLRPD 8278
QY 6 ----- 5
Db 8279 HGLYMKVENDHGIAPACTSVLDTPGPPIINFVEDIRKTSVLCKWPEPLDGGSEIIN 8338
QY 6 ----- 5
Db 8339 YTEKKDKTPESEWIVVSTLHCKYSVTKLIEGKEYELFRVAENRFGPPCVSKPLV 8398
QY 6 ----- 5
Db 8399 AKDPGPDPADPKPIVEDYTSNMLVKWNEPKDNGSPILGYWLEKREVNSTHMSRVNKL 8458
QY 6 ----- 5
Db 8459 LNALKANVDGLLEGLTYFVRVCAENNAAGPKESPSPDPTAHDPISPGPIPRVTDSS 8518
QY 6 ----- 5
Db 8519 TTELEMEPPAFNGGEIYGFVDKQVGNKMSRCTEMIKVROYTVVEIREGADYKLR 8578
QY 6 ----- 5
Db 8579 VSAVNAAGBGPGETQPVTVAEPOBPAPVELDVSKGIQIMAGKTLRIPAVVTGPRVPT 8638
QY 6 ----- 5
Db 8639 KWTKEEGELDKDRVIDNVGTSKSELLIKDALRKDGRVYITATNSCGSFAARVAVED 8698
QY 6 ----- 5
Db 8699 VPBPVLDLKPVTNRMKCLNNSDPEDDGSSEITGFIIEKKAAMHTWQPIETERSKCD 8758
QY 6 ----- 5
Db 8759 ITGLEGOYKFRVIKAKNKGCPVEYELPILAVDPLGPTSPERLTYTEROSTITLDM 8818
QY 6 ----- 5
Db 8819 KEPRNSGSPIOGYIIEKRHRDKPDERVKNKRLCTPTSLVENLDEHOMYEFVRKAVMEI 8878

QY 6 ----- 5
Db 8879 GESEPSLPLNVVIODEVPPTIKRLSVNGDTIKVAGBPVHIIPADYVGLMPKIEMSKN 8938
QY 6 ----- 5
Db 8939 ETVIEKPTDALQITKEVSRSEAKTELIPKAVREBKGYVTWASNRLGSVRNVHEVY 8998
QY 6 ----- 5
Db 8999 DRSPPRNLAVTDIKAESCYLTWDAPLNDNGSEITHYIDKRDASRKAEMEYTNNAVE 9058
QY 6 ----- 5
Db 9059 KRYGIWKILIPNGYEFVRVAVNKYGISDECKSDKVVYIDPYRLPGPPGKPVYARTKSM 9118
QY 6 ----- 5
Db 9119 LVSWTPPLDNGSPITGYWLEKREBGSPLYWSRSHAPITKVOLKGEFNVPLLEGVYQ 9178
QY 6 ----- 5
Db 9179 FRAMAINAGIGPSPBSPDEVAGDPIFPGPSPCEVKDKTKSISLGMPKPAKDGSP 9238
QY 6 ----- 5
Db 9239 INGVIEMOEGSTDMKRVNEPDKLITCEGVPLKELRKYRFRVAVNAGESEPSDT 9298
QY 6 ----- 5
Db 9299 TGEIATDIOEBEYFIDIGADCLVCKAGSOIRIPAVIKGRPPKSSWEDGRKAKAMK 9358
QY 6 ----- 5
Db 9359 DGVHDIPEDAQLETAENSSVIIIECKRSHGKYSITAKNAGOKTANCRKAVADVGP 9418
QY 6 ----- 5
Db 9419 KDLKVDITRGSCLSMKMPDDGDRIGYIEKRTIDGKAMTKVNDGSGTTFVVDL 9478
QY 6 ----- 5
Db 9479 LSEQYFFRVRAENRFGIPPEYETIQTATADPIYPPDPPIKLIGLITKNVHL SMKPP 9538
QY 6 ----- 5
Db 9539 KNDGSPVTHYIVECLAMPDTGTTKEAMRQCNRDVEELOFTVEDLYBGEYEFVRKAVN 9598
QY 6 ----- 5
Db 9599 AAGVSKPSATVGPCDCORPDPSPSIDLKEFMEVEEGTNVNIIVAKIKGVPEFTLWFKAP 9658
QY 6 ----- 5
Db 9659 KRPDKPEVLYDTHVKNKLVYDTCILVYIPOSRSBDGLYTTAVANNLCTASKEMRLNVLG 9718
QY 6 ----- 5
Db 9719 RPPVGPVIFKESYADQMTLSWFPKDDGSKITNVYIEKREANRKTWVHVSSEPECT 9778
QY 6 ----- 5
Db 9779 YTIPLLEGHEYVRIMANQNTYIGEPDSEPTARNLFSVPAGDPKPTVSVTSRNSMTV 9838
QY 6 ----- 5
Db 9839 NMBEPYDGSPLYGYWLEMKDITSKRMKRVNRDPIAMTLGVSYKVTGLIEGSDIOFRV 9898
QY 6 ----- 5
Db 9899 YAINAGVPASLSPDATABDPIAPGPFPKVTWTKSSADLEWSPPLKDGSKVTGY 9958
QY 6 ----- 5

Db 9959 IVEYKEGKEEMKGDKEVGRKTLVYTGKLGCAFYKFRVSAVNAGIGEPGEVTVIEM 10018
 QY 6 ----- 5
 Db 10019 KDRUVSPDLQLDASVDRIVVHAGVIRIAYSGKRPPTVWNNMERLPOEATJETTA 10078
 QY 6 ----- 5
 Db 10079 ISSSMVICKNCQSHQGVYSLAANKNEGERRKTLIVDLVPGFVGPPLAHLNLTNESCKL 10138
 QY 6 ----- 5
 Db 10139 TWSPEDDGGSPITNVIYEKRESDRAMPVYTVTRQNTVGLIQKAFYFRIAENS 10198
 QY 6 ----- 5
 Db 10199 IGMGPEVETSEALVIREPITVPERPEDLEVKEVTKNTVTLTNMPKPYDGGSEIINYLES 10258
 QY 6 ----- 5
 Db 10259 RLIGTEFHHVYNDNLISRYTVYKLGKEDPYEYRVSAVNIYGQKPSCTKPICTKDEL 10318
 QY 6 ----- 5
 Db 10319 APPLHLDFRDKLIRYGEAFALGRYSKPKPKVSWFKDEADVLEDDRTIKTPATLA 10378
 QY 6 ----- 5
 Db 10379 LEKIKARSDSGKYCVVENSTGSRKGCQVNVVDHBPVPGVSPFDEYTKDYMVISMKP 10438
 QY 6 ----- 5
 Db 10439 PLDDGSKITNYIIEKKEVGKDWMPVTSASAKTCTKVSLLBGKDYIFRIHAENLXGIS 10498
 QY 6 ----- 5
 Db 10499 DPLVSDSMKAKDRPRVPDAPDQPIVETKDSALVTWNNKPHDGKPIYNIIEKRETMK 10558
 QY 6 ----- 5
 Db 10559 RMAVATDPIHPYTKFRVPLLEBGCQYEFVSAENEGIGDPSPPSKPVAKDPIAKPSP 10618
 QY 6 ----- 5
 Db 10619 PVNEAIDITCNISVDLTWOPRPHDGSKILGIYEQKVGDEEMRANHTPESOPETKXK 10678
 QY 6 ----- 5
 Db 10679 VTGLRDCQYKFRVLAVNAGESDPAHVPEVLVKRLEPPELILDANMARQOHIKVDT 10738
 QY 6 ----- 5
 Db 10739 LRLSAILKGVPPKVTWKEDRDAPYKARIDVTPVSGKLEIRNAHEDGGLSYLVENPA 10798
 QY 6 ----- 5
 Db 10799 GSKTVSVKVLIDKPPROLEVEIRKDCYLFWKEPLDDGGSVITNYVERADVSAQ 10858
 QY 6 ----- 5
 Db 10859 WSPLSATSKKSHPAKHLNEGNOYLFRVAENOGYGRGFVETPKPIKALDPLHPPGPKD 10918
 QY 6 ----- 5
 Db 10919 LHHVDVDTESVLYWNNKPRDGGSPITGIYVEYQEBGTQDMIKKTYTNLECVYTGLOQG 10978
 QY 6 ----- 5
 Db 10979 KTYRFRVKAENIVGIGLBDTIIPIECQEKLVPSVELDVKLEGLVYKAGTVAFPALIR 11038
 QY 6 ----- 5

Db 11039 GVPVPTAKWTIDGSEIKTDEHYVETDNFSSVLTIKNCILRHDGTGEYOITVSNAGSKTVA 11098
 QY 6 ----- 5
 Db 11099 VHLTVLDVPGPPTGPIINILDVTPHEMTISWOPKDDGGSPIYINIVKODTRKDTGCVS 11158
 QY 6 ----- 5
 Db 11159 SGSSKTKIKIPHLQKCEYFRVAENKIGVPLDSTPIYAKHKFSPPSPGKPYVTDI 11218
 QY 6 ----- 5
 Db 11219 TENATVSWTLPKSDGSPITGIYMERREYTGKWRVWKPPIADLKFRVYGLYEGNTYEF 11278
 QY 6 ----- 5
 Db 11279 RVFAENLAGLSPSPSPDIACRPIKPPPIPNKLDKSRRETADLVWTKPLSDGSP 11338
 QY 6 ----- 5
 Db 11339 LGYVECOKPGTAOMNRINKDELIRQCAFVYVGLIEGNEYFRRIKANIYEGEPRELA 11398
 QY 6 ----- 5
 Db 11399 SYIANDILHPPEVELDYTCRDVITVRGQITRILARVGRPEPDIWTKEGVLVREKRV 11458
 QY 6 ----- 5
 Db 11459 DLIQDLPRVELQIKAVRADHGKTIISAKNSSGHAQSAIYVNLDRPQCNLKVNTWK 11518
 QY 6 ----- 5
 Db 11519 ENCTISWENPLDNGGSEITNFIYERKPNQKGSIVASDVTKRLIKANLANNEYFRVC 11578
 QY 6 ----- 5
 Db 11579 AENKVGSPITETKPIIAPNIDRGPENILHADKGTVYIKWRPDDYDGGSPNLSY 11638
 QY 6 ----- 5
 Db 11639 HYERRLKGSDBMERVHKSGIKETHYVDCVNOYIEFRVQTKNKGSGSDWVKEEYVVK 11698
 QY 6 ----- 5
 Db 11699 EDLQKRVLDLKLGLVLTAKAGDTIRLEAGVGRKPPPEVAMTKDKDADTLTRSPRYKIDTR 11758
 QY 6 ----- 5
 Db 11759 ADSKFSLTAKRSDGKIVTATNTAGSFVAIYATVNLKPGVYRNLKIYDVSSDRCTV 11818
 QY 6 ----- 5
 Db 11819 CWDPEDDGCELONYIIEKCTFKRWVSTYSATVLTPTQTVTRILIENEXYFRVAENK 11878
 QY 6 ----- 5
 Db 11879 IGTPTESKPYIAKTKYDKPGRDPPEVTYKSKEMTVWNNPBYDGKSITGYFILEKK 11938
 QY 6 ----- 5
 Db 11939 EKHSTRWVYVNSALPERMAKQONLLPDHEYQFRKAENEIGISPLSPRYVAKDPIE 11998
 QY 6 ----- 5
 Db 11999 PPGPTNFRVYDTTKHSITLGMKPYDGGAPLIGYVEMRPIADASPDDEGKRCNAAA 12058
 QY 6 ----- 5
 Db 12059 QLVKKEFTVSLDENQYEFVCAONQYIGRPAELKAIRKEIIEPPEIDLASMKRL 12118
 QY 6 ----- 5
 Db 12119 VIVRACPIRLFAIYGRAPAYTWKRGIDNVVRKGOVDLVDTMAFLYINSTDDSGK 12178

```

QY 6 ----- 5
Db 12179 YSLTLVNPAGEKAVFNVAVLDPGFVSDLKYSVDTKTSCHVSWAPENDGGSYTHYIV 12238
QY 6 ----- 5
Db 12239 EKREADRKTWSTVTPPEVKTSFHHVNLVPGNHYFRVTAVNEYPGVPPTDVPKPVLASDP 12298
QY 6 ----- 5
Db 12299 LSEDPDRKLEATEMTKNSATLAWLPLBLDGAKTIDGYIISYREBQPADRWTESYVKD 12358
QY 6 ----- 5
Db 12359 LSLVYTGLEKGRKRYKFRVARNVAVSLPREAGVEAKEQOLLPPKILMEQITIKAGK 12418
QY 6 ----- 5
Db 12419 LKLEAHVYGRPHPTCKMKKGEDEVYTSHLAVHKADSSILIKDVTKRKDSGYSLTAEN 12478
QY 6 ----- 5
Db 12479 SSGTDTQIKVYVMDAPGPPDPIDSIDADAGLSMHIPLDGGSNITNYIVEKDV 12538
QY 6 ----- 5
Db 12539 RGDWYATALASVTKTSCRWGKLIPGOEYIFRVAENRFGISEPLTSPKVAQFPFGVSEP 12598
QY 6 ----- 5
Db 12599 KNAHYTKVKKDCIFVAMDRPDSGGSPILIGYILIEKKENSLMVKANDTLVRYSTEYPCAG 12658
QY 6 ----- 5
Db 12659 IVEGLEYSFRYIALNKAAGSSPPSKPTLEYVYAMVDPGKPEVIDVTKSTVSLIWARBK 12718
QY 6 ----- 5
Db 12719 DGGSTIIGYFVACKLPDGDKNVRCNTAPHQIPQEEYATGLEKAQYFRAIARTANIS 12778
QY 6 ----- 5
Db 12779 PPSESDPYTLAENVPPRIDSVAMKSLTYKAGTNYCLDNTVGRKMPYVSMKKDGL 12838
QY 6 ----- 5
Db 12839 LKPBGIKMAQONLCTLEFSVNRKDSGDYITTAENSSGSKSATIKLVLDPGPASV 12898
QY 6 ----- 5
Db 12899 KINKMYSDBRAMLSWEPPLDGGSEITNTITDKRETSRPNMAQVSATVPITSCSVEKLLIEG 12958
QY 6 ----- 5
Db 12959 HEYQFICAENKYGVDVTEPEALAKNPDPGRCDPVVISNITTKDHMTVSMKPPADG 13018
QY 6 ----- 5
Db 13019 GSPITGYLLEKRETOAVNMKTVNRKPIIERLTKATGLOEGTEYEFRTVAINKAGPGKPSD 13078
QY 6 ----- 5
Db 13079 ASKAAVADPOVPPAPPAFPKYVYDITRBSVLSLWKGPAVDGSPITIGLYLEVKKRADSNN 13138
QY 6 ----- 5
Db 13139 VMCNLPQNLQKTREVTGLMEDTQYQFRVAVANKIGYSDPSOVDPKHVPKDLIPPEGEH 13198
QY 6 ----- 5
Db 13199 DADLKKTLILRAGVTMRLYVPVKGRRPPKTTWSKPNVMLRDRIGLIDIKSTDFLRCEN 13258

```

```

QY 6 ----- 5
Db 13259 VNKYDAGKYLITLBNSSGKREYTIIVKVLDPGPPINVTYKEISKDSAYTWEPPIIDG 13318
QY 6 ----- 5
Db 13319 SPIINYVQKDAERKSWSTVTTECSKTSFVRPNLIEGKSYFFRFAENEGIGDPQETR 13378
QY 6 ----- 5
Db 13379 DAVKASQTPGVVDLKYRSVSKSSCSIGMKKPHSDGSRITIGYVDFLTEENKQRYMKS 13438
QY 6 ----- 5
Db 13439 LSLQYSAADLTEGREYTFRVAENENEGTPESEITTVARDDVAPDLDLKGPDLCYAK 13498
QY 6 ----- 5
Db 13499 ENSNFRLLKIPKRPAPSVSMKKGEDPLADTVRSVSSAVNTLLIYDCQKSDAGKTTI 13558
QY 6 ----- 5
Db 13559 TLKNVAGTKEGTISIKVYGRGPIPTGPDKFDEVTAEAMTLKMAPPKDDGSEITNTILEK 13618
QY 6 ----- 5
Db 13619 RDSVNNKWTVCASAVQKTFEVRTRLHEGMEYTFRVAENKYGGEGLKSEPIYARHPDV 13678
QY 6 ----- 5
Db 13679 PDAPPPNIVDVRHDSVSLWTDPKKTGSPITGYHLEFRKNSLMKRAKTPIRMRDF 13738
QY 6 ----- 5
Db 13739 KYTGLEGEYFRVMAINLAGVKSPLSEPVVADPIDPQKPEYINITRNSVTLIMT 13798
QY 6 ----- 5
Db 13799 EPKYDGHKLGYLVERKDLPSKSMKANHVNPECATVTLDEVGKYEFRIRAKNTAG 13858
QY 6 ----- 5
Db 13859 AISAPESTETIICKDEYEAFTVLDPTIKDGLTKAGDTIVINAISILGKPLPKSWSK 13918
QY 6 ----- 5
Db 13919 AGKDIRPSDITQITSPTSSMLTIKATRKDAGEYITITATNPSTKVEHAKTVLDPG 13978
QY 6 ----- 5
Db 13979 PGVYEISVSAEKATLWTPPLEDGGSPISKYILIEKRETSRLMTVSEDIQSCRHVAIK 14038
QY 6 ----- 5
Db 14039 LIQONEYIFRVSANVNHGKEPQOSEPVKAVDRPGPEKEVSNVTKNATVSMKRP 14098
QY 6 ----- 5
Db 14099 VDDGSEITGYHERREKSLRWVRAIKTPVSDLRCKVYGLQEGSTYEFRRVSAENRAGIG 14158
QY 6 ----- 5
Db 14159 PPEASDSVLMKDAVPPGPPSNPHVDTTKKASIASLWGPBHYDGLLEITGYVVEHQYK 14218
QY 6 ----- 5
Db 14219 DEAMIKDTGTALRITQFVVPDQTKENYFRISALINDAGVGPANVIPDVEIYERMAAD 14278
QY 6 ----- 5
Db 14279 FELDAELRRLVVRAGLSIRIFVPIKGRPAEYTWTKDNINLKRANIENTESFTLLIIP 14338
QY 6 ----- 5

```

Db 14339 EGNRYDYGKFWMTLENPAGKSGFVNVNVRVLDTPGPVLTLPDITKDSVTLHMDLPLIDG 14398
QY 6 ----- 5
Db 14399 GSRITNYIYERKREATRKSYSTATTCKHCYKVTGLSEGEYFVRMAENEXGIGERTET 14458
QY 6 ----- 5
Db 14459 TEPEKASEAPSPDSLNMIDITSYVSLAMPKPKHDGSKITGYVLEAQKSGSDQTHIT 14518
QY 6 ----- 5
Db 14519 TVAGLECYVNLTEGEERTFQVMAVNSAGSAPRESHPVLEQTMLELDLGIYQKLV 14578
QY 6 ----- 5
Db 14579 IAKAGDNKVEIPVLRPKPTVTKKGDQILKOTQRVNFTTATSTLLINECVRSDDGP 14638
QY 6 ----- 5
Db 14639 YPLTARNIVEGVDVITIOYHDIKPGPTGPIKFDVSSDVFTEWMDPPENDGVPISNTV 14698
QY 6 ----- 5
Db 14699 VEMKQDSTTWVIELATVIRTYKATRLTGLYOFVRKAQNRVYVGPQITSAMIVANP 14758
QY 6 ----- 5
Db 14759 FKVPGPPTQOVAVTSDSMTISWHEPLSDGSPILGTYHKKERNGILMQYVSKALVYG 14818
QY 6 ----- 5
Db 14819 NIFKSSGLTDGIAVEFRVIAENNAKSKPSKPEPMALDIPDPGKVPVLTNRHTVTL 14878
QY 6 ----- 5
Db 14879 KWAKPEYTGFKIYSYIYERKDLPNGRMLKANFSNILENFTVSGLTEDAAYEFVYIAKN 14938
QY 6 ----- 5
Db 14939 AAGAISSPSEPSDAITCRDVEAPKIKVDYKFKDVTILKAGEAFLEADVSGRPPIMEM 14998
QY 6 ----- 5
Db 14999 SKDGELEGTAKLEIKIADESTLNVNKRSDRSGAYTLTATNPGGFAKHIFNVKVLDRP 15058
QY 6 ----- 5
Db 15059 GPPGPIAVTEVTSEKCVLSWFPPLDDGAKIDHYIVOKRETSRLANTVNAVSEVOYTKLK 15118
QY 6 ----- 5
Db 15119 VTLLKNGEYIFRVMAVNNKGVGEPLSEPVLANPYGPPDPKPNPEVTTITKDSMWYCW 15178
QY 6 ----- 5
Db 15179 GHPDSDGSEIINYVERBDRKAGQRIKCNKTLTDLRYVSGLTGEGHEFRIMAENNA 15238
QY 6 ----- 5
Db 15239 GISAPSPSPRYKADTVFKRPPGNRVLDTSRSSISIAMNKP IYDGGSEITGYMVEIA 15298
QY 6 ----- 5
Db 15299 LPREDEMOTYTPAGLAKATSTTITGLTENDEYKIRIYAMNSEGLGEPALVPOTPKADRM 15358
QY 6 ----- 5
Db 15359 LPPEILDADLRKVITIRACCTLRLEFVPIKGRPDEYKWARDHGESIDKASIESASSYTL 15418
QY 6 ----- 5

Db 15419 LIVGNVNRDPSGKYLITVENSSGSKSAFVNVNVRVLDTPGPQDLKVKKEVTKUSTVLTWDP 15478
QY 6 ----- 5
Db 15479 LIDGSKIKNTIYERKRESTRATVATNCHKTSMKYDQLOEGCYFRVLAENEGIGL 15538
QY 6 ----- 5
Db 15539 PAETAESYKASERPLPCKITLMDYTRNSVSLSMKEPEHDGSRLLGYIVEMQTKGSDKW 15598
QY 6 ----- 5
Db 15599 ATCATVKTATITGLIGEEYSFRVSAQNEKGISDPRQLSVPIYANDVYIPAFKLLFN 15658
QY 6 ----- 5
Db 15659 TPTVLAGEDLVDPFVIGRPTPAVTHKDNVPLKQTTVYNAESTENNSLTIKACREDV 15718
QY 6 ----- 5
Db 15719 GHYVVKLINSAGEALETNLNVIYLDKPPGPTGPKMDEVLTADITLSWGPFXDGGSSINN 15778
QY 6 ----- 5
Db 15779 YIVKRDJSTTWOIVSATVAFITIKACRLTGCYOFRIAEENRYCKSTYLNSEPTVAQ 15838
QY 6 ----- 5
Db 15839 YPFKVPGPPTPVYVTLSSRDSMEVQWNEPIDGSGRVIGHLEKERNISLIMVNLNTP 15898
QY 6 ----- 5
Db 15899 PQTKFKTGLBEGVEYEFVSAENIVGIGKPSKSECVARDPCDPGRPAIIVTRNSV 15958
QY 6 ----- 5
Db 15959 TLOMKKPYDGGSKITGYIYVEKELPBGKMKASFTNIIDHFEVYGLVEDHREFRVIA 16018
QY 6 ----- 5
Db 16019 RNAAGVFSEPSSTGALTARDEVDPRIKMDPKYDITVHAGESFKVDADYKPIPTI 16078
QY 6 ----- 5
Db 16079 QWIKGDELSTARLEIKSTDFATSLSVKDAVRVDSGNYIILKANVAGERSVTNVKVL 16138
QY 6 ----- 5
Db 16139 RCPPEGPVVISGVTAEKCTLAMKPPLODGGSDIINTYVERRETSRLVTVVYDANVOTLS 16198
QY 6 ----- 5
Db 16199 CVTKLLBEGNEYFRIMAVNNKYVGEPLSEPVAKNPFVVDAPKAEVTTVKDSMIV 16258
QY 6 ----- 5
Db 16259 VNERPASDGSSEILGYLEKRDKEGIRMTCHKRLIGELRLRYTGLLENHDEFRVSAEN 16318
QY 6 ----- 5
Db 16319 AAGLSRPPSPAYOKACDPIYKPPNNPKVIDITRSSVFLSMKPIYDGGCEIUGYIVE 16378
QY 6 ----- 5
Db 16379 KCDVNVGEWIMCTPPLGINKTNIYEKELLEKHEYNFRICAINKAGVGEHADVPGLIVE 16438
QY 6 ----- 5
Db 16439 KLEAPDIDLELRKIINIRAGSLRLFVPIKGRPTPEWKGVKDGELRDAIIDVTSSE 16498
QY 6 ----- 5
Db 16499 TSLVLDNVRYSKTYTLTLENSSGTKSAFYTVVRVLDTPSPPVNLKVEITKDSYSTIWE 16558


```

QY 6 ----- 5
Db 16559 PPLIDGSKIKNIYVEKREATRKYAAVVTNCHKNWKIDQJGCGSYFPAENYGI 16618
QY 6 ----- 5
Db 16619 GLPAQOTADPIVAEVPQPGKITVDVTRNSVLSWTRPEHGGSKIIQYIVEMQAKHSE 16678
QY 6 ----- 5
Db 16679 KWSECARXSLQAVITNLJGSEELFRVAVANEKGRSDPSLAVPIYAKDLVIEPDYKPA 16738
QY 6 ----- 5
Db 16739 FSSYSVOVGODLKMEDPISGRPKPTITWTKDGLPKQOTRINWDSLDLTLSIKETHKD 16798
QY 6 ----- 5
Db 16799 DGGQGITVANVGOCTASIEITVLDKPDPPKGFVKFDVSAESTILSMNPLYTGCCQI 16858
QY 6 ----- 5
Db 16859 TNYIVOKRDTTIVWDVVSATVARTILKYTKKTGTGYOFIRFAENRYGSPALSDPIY 16918
QY 6 ----- 5
Db 16919 AQYPKPEPGPTEPFAAISKDSMVIQMHBPVNNGSPVIGYHLERKERNIIIMTKVNT 16978
QY 6 ----- 5
Db 16979 IIHDTQKQNLGEIEYERFVAVENIVGVKASKNSECYVARDPCDPPTPEPIWKNRN 17038
QY 6 ----- 5
Db 17039 EITLQWTRPYVDGSGMITGYIYERKDLPGRWKMASTNVLETQFVSGLTEDQRYEFY 17098
QY 6 ----- 5
Db 17099 IAKNAAGAIKSPDSSTGPTIAKDEVELLPRISMDPKFRDTIVNAGETFRLEADVHGKPLP 17158
QY 6 ----- 5
Db 17159 TIEMLRGDKIEESARCEIKNTDFKALLIYKDAIRIDGQYIIRASNVAGSKSPVAVKY 17218
QY 6 ----- 5
Db 17219 LDRGPPEGVQYVGTSEKSLTWSPPLQDGGSDISHYVERKETSRLAMTVVASEVYT 17278
QY 6 ----- 5
Db 17279 NSLKYTKLEBNEFYRIMAVNKYGVGEPLSAPVLMKNPFVLPDPKSLLEVNTIAKDSM 17338
QY 6 ----- 5
Db 17339 TYCWNRPDSGSEIIGYIYERKDRSGIRMIKCNKRITDLRLRVYGLTEDEHYEFYVSA 17398
QY 6 ----- 5
Db 17399 ENNAGVEPSPATVYKACDPVEKPGPPTNAHIVDTKNSITLAMGKPIYDGSEIIGYV 17458
QY 6 ----- 5
Db 17459 VEICKADEEMQIVTPOGTGLRTRFELSKLTHEQYKIRCALNKVGLGSAVSVPQTVAP 17518
QY 6 ----- 5
Db 17519 EDKLEAPELDLDSELKRGIYVAGSARIHIIPKGRPMPEITWRSREGFTDKVQIEKGV 17578
QY 6 ----- 5
Db 17579 NTYQLSINDCRNDAGKIYIILKENSSGSKSAFTVAVLDTPGPPONLAVKVRKDSAFIV 17638

```

```

QY 6 ----- 5
Db 17639 WEPIIDGGAKVKNYVIDRSTRKAYANVSSKCSKTSFKVENLEGALIYFRMAENEF 17698
QY 6 ----- 5
Db 17699 GVGVPVETDVAKAAPPSPPGKVTLLDVQSOTASIMWEKPEHGGSRVIGYVENQPKG 17758
QY 6 ----- 5
Db 17759 TEKWSIVAESKVCNAVVTGLSSGOEYQFRVAKYNEKKSPPVGLGVYIANKDLTTOPSLK 17818
QY 6 ----- 5
Db 17819 LPENTYSIQAGBDLKIPIVIGRPNISWVMDGEPLKQOTRVNVEBATSTVLHKBGN 17878
QY 6 ----- 5
Db 17879 KDDFGKYTVATANSAGTATENLSVILEKPPVGPVREDEVSADPVVISMEPAYTGCC 17938
QY 6 ----- 5
Db 17939 QISNTIYERKDTTTTWHVVSATVARTTIKTKLTGTGYOFIRFAENRYGKSADLSKA 17998
QY 6 ----- 5
Db 17999 VIVQYFKPEPGPTEPFAAISKDSMVIQMHBPVNDGCKIIGYHLQEKERNIIIMTKLN 18058
QY 6 ----- 5
Db 18059 KTEPIDTKERTGLDEGLEEFKVSANIVIGIKPSKSECFVARDPCDPGRPEAIYIT 18118
QY 6 ----- 5
Db 18119 RNNVTLMKKRPAYDGSKITGYIYERKDLPGRWKMASTNVLETQFVSGLTEDQRYEF 18178
QY 6 ----- 5
Db 18179 RVIARNAAGNFSEPSDSSGAITARDEIDAPNASLDPKYKDVIVHAGTFVLEADIRKRP 18238
QY 6 ----- 5
Db 18239 IPDVWMSKDGKLEETARMEIKSTIQKTLVVKDCIRTDGQYILKLSNVGTSKIPIT 18298
QY 6 ----- 5
Db 18299 VKVLDRCGSPBGLKVTGVTAEKCYLANNPPLQDGCANISHYIIEKRETSRLSWTQVSTE 18358
QY 6 ----- 5
Db 18359 VOALNKYTKLPGNEIYFRVAVANKYIGEBLESQPYACNPKYKPPGSPNPEVSATIK 18418
QY 6 ----- 5
Db 18419 DSMVYTWAPVVDGGTEIEGTILEKRDKEGVKWKCNKKTLLDLRLRVYGLTEGHSEYFR 18478
QY 12 ----- 11
Db 18479 VAENNAAGVEPSEPSVYRACDALYPGPPSPNPKVTVTSRSSVSLAMSKPIYDGAAPVK 18538
QY 12 ----- 11
Db 18539 GYVEVEKAADDEWTTGPPGLQCKOPTVYTKLKNTEYNFRCALINSEGVGEBATILPGS 18598
QY 12 ----- 11
Db 18599 VVAQERIEPEIELDADLRKVVVLRASATLRLFTYIKGRPEYKWEKAESILTDRAOIE 18658
QY 12 ----- 11
Db 18659 VTSFTMLVINDVTRFDSGRVNLLENNGSKTAFVNVRLDSPAPVNLTRREVKDSV 18718
QY 12 ----- 11

```

Db 18719 TLSWEPPLIOGAKITNVIYEKRETTKRAVATITNNCTKTTFRIENQEGCSYFRIAS 18778
QY 12 ----- 11
Db 18779 NEYIGLPAETTEPVKSEPLPPGRVTLVDVFNATITKWEKESDGSKITGYVEMQ 18838
QY 12 ----- 11
Db 18839 TKGSEKSTCTQVKTLEATISGLTAGEEYFVRAAVNEKGRSDPROLGVPVIARDIEIKP 18898
QY 12 ----- 11
Db 18899 SVELPHTFNKAREQKIDVPFKGPOATVNMKDGQTLKETTRVNVSSKTYLSLK 18958
QY 12 ----- 11
Db 18959 EASKEDVGTVELCVSNSAGSITVPITILDRPGPPRIDEVSCDSTITSMNPEYDG 19018
QY 12 ----- 11
Db 19019 GCQISNIVYEKRETTSTMTHTVSOAVARTSIIKIVRLTTSGETOPRVCANRYGKSSSES 19078
QY 12 ----- 11
Db 19079 SAVVAEYFPSPGPPGPKVHATKSTMLVTWQVPVNDGSGRVIGYHLEKERSIIMSK 19138
QY 12 ----- 11
Db 19139 ANKILIDTQVKSGLDEGLMTEYRTAENIAGIKCSKSCPEVPANDPCDPGQPEVTN 19198
QY 12 ----- 11
Db 19199 ITRKSVLSKMSKPHYDGAKITGYIVERRELPGRMWKCNVTNIQETTYFEVTELTEDQRY 19258
QY 12 ----- 11
Db 19259 EFRVAFARNADSVSESESTGPITIKDVEPPRVMDVKFRDVIYVAGEVLIKINADIAG 19318
QY 12 ----- 11
Db 19319 RPLVYSNAKDGIELEEARTEIISTDNHILLTVKDCIRDTGQVLTILKNVAGTSVAV 19378
QY 12 ----- 11
Db 19379 NCKVLDRKPGPAGPLEINGLFAEKCSLSWGRPOEDGADIDYHRRKRETSHLAMTICEG 19438
QY 12 ----- 11
Db 19439 ELQMTSCRYTKLKGNEYIFRVTGVNKYVGEPLESAIKALDPFTVPSPTSLEITSVT 19498
QY 12 ----- 11
Db 19499 KESMTLCMSRPESDGSISGIIIRREKNSLRWVRVKKRYVYLKXSTGLRGCEY 19558
QY 12 ----- 11
Db 19559 RYVANAAGLSLPSETSPILRAEDVPFLPSPSKPIVDGKTTITIAWVPLFDGAPI 19618
QY 12 ----- 11
Db 19619 TGYIYEKKSDDTDWKTISIQLRGTEYISGLTTGAETVFRVKSVMKVGASDPSDSDPO 19678
QY 12 ----- 11
Db 19699 IAKEREERPLDIDSEMKITLIVAGASFTMTVPRGRPVNVLMKRPDITLRTRAYDT 19738
QY 12 ----- 11
Db 19739 TDSRTSLTIENANRNDGKTYTLIQNVLSAASLFLVAVLDPGPPNITVQDVKESAV 19798
QY 12 ----- 11

Db 19799 LSWDVPENDGAPVKNYHIEKREASKKAWSVTNNCNRLSYKVTNLOGAITYFVSGEN 19858
QY 12 ----- 11
Db 19859 EFGVGIPAEFKGKGVKITTEKPSPEKLVTSISKDSVSLTWLAKPEHDSGRIVHYVEALE 19918
QY 12 ----- 11
Db 19919 KQKMMVKCAVAKSTHHVYVSGLRNSEYFVRFAENQAGLSDPRELLPLVLIKQLEPPE 19978
QY 12 ----- 11
Db 19979 IDMKNEPSTHYVRASNLKVDIPISGRPLPKVTLSDRGVPLKATMRENTETIAENLTIN 20038
QY 12 ----- 11
Db 20039 LKESVTADAGREITRANSSTTKAPINIVLDRPGPPGVPVISDITEESVTLKWEPPK 20098
QY 12 ----- 11
Db 20099 YDGSQVNTYILLKRETSVAVTEVSATVARTMKVMKLTGEEYQFRKAENRPGISDH 20158
QY 12 ----- 11
Db 20159 IDSACVTRLPYTPPGPSTPMVTNVTRESTIVGHEPVNSGSAVGYHLEMDRNSIL 20218
QY 12 ----- 11
Db 20219 WOKANKLVIKTHFKVYITISAGLIYEFVYAEANAGVGRPSHPSEVLAIDACEPRRVR 20278
QY 12 ----- 11
Db 20279 ITDISKNSVLSWQOPAFDGSKITGYIVERDLDPGRWTKASFNTVETOTISGLTON 20338
QY 12 ----- 11
Db 20339 SOYEFVAFARNAGSISNPEVVGPICTIDSYGVPYIDPLEYTEYVYKTRAGTSYKLNAG 20398
QY 12 ----- 11
Db 20399 ISGRPAITIEWYKDKELQTNALVCVENTDILASILIKDADRLNSGYELKRNAMASAS 20458
QY 12 ----- 11
Db 20459 ATIRVOILDKPGPGPIEFKVTAEKITLMLRPPADGAKITHYIEKRETSRVVSM 20518
QY 12 ----- 11
Db 20519 VSEHLECIITTTIKITKNEYIFRVAANKYIGIPELSDSVYAKNAFTVPGPGIPEVT 20578
QY 12 ----- 11
Db 20579 KITKNSMTVWNSRPIADGSDISGYFLERDKKSLGMFVLKETIRDFRQKVTGLTENS 20638
QY 12 ----- 11
Db 20639 YQYVCAVNAAGOPSESEPEYKADPIDPGPPAKIRIADSTKSSITLIGMSKPYDGG 20698
QY 12 ----- 11
Db 20699 SAVTYVEIROGEFEEMTTVSTKGEVKTTEYVSNLKGVMYFRVSAVNCAGGEPTE 20758
QY 12 ----- 11
Db 20759 MNEPVQAKDILEAPEIDLVALRTSVIAKAGEDVQVILPFKGRPPPTVWRDEKNIGSD 20818
QY 12 ----- 11
Db 20819 ARYSIENTDSSLLTIPQVTRNDTGKYLITENGVEPKSSTVSVKVDTPAACOKLOVK 20878
QY 12 ----- 11
Db 20879 HVSRTVTLMDPPLIDGSPITINVIYEKRDATKRTVSVSHKCSSTFKLIDSEKTPF 20938

```
QY 12 ----- 11
Db 20939 FFRVLAENEIGIGEPCTETPEPKAAEPAPIRDLNMDSTKTSVILSWTKPEDGGSVIT 20998
QY 12 ----- 11
Db 20999 EYVVERKKGGEQWWSHAGISKICEIVSOLKEQSVLEFRVFAKNEKGLSDPVTIGPIYVK 21058
QY 12 ----- 11
Db 21059 ELITTEVDLSDIPGAQVTVRIGHNVHLELPYKGRKPSISMLKQGLPLKESFVRFSKT 21118
QY 12 ----- 11
Db 21119 ENKITLSINAKKEHGKTYTILDNACRIAVPIVITLGPSPKPKPIRDEIKADSVI 21178
QY 12 ----- 11
Db 21179 LSWDPVEDNGGGEITCYSTIEKRETSQTNKMKVCSVAFTTFKVPNLVKAETQFRVRAEN 21238
QY 12 ----- 11
Db 21239 RGVQSOPVSSIIIVAKHOFIRBPGRPKVITYNVTSDDMSLITWDAPYDGGSEVTFHVER 21298
QY 12 ----- 11
Db 21299 KERNSILMOKVNTSPISGREYRATGLVEGLDQFRVYAENSAGLSSPDSKFTLAVSPV 21358
QY 12 ---TPD----- 14
Db 21359 DPGGFPDYIDVYRETTLKMPPLRDGSKIVGYSIEKRGNERVRCNFTVSECQITV 21418
QY 15 ----- 14
Db 21419 TGLSPDRYEFRIANNAVGTISPPSSGIIIMTRDENVPPIYEGPEYFDGLIISGHS 21478
QY 15 ----- 14
Db 21479 LRIKALVOGRPVRYTWKFDGVEIEKRAMEITNVLGSTSLVRDATTRDHRGVYVEAKN 21538
QY 15 ----- 14
Db 21539 ASGSAAEIKVAVODTPGKVGPIRFTNITGEKMTLMDAPLNDGCAPITHYIIEKRETS 21598
QY 15 ----- 14
Db 21599 KIAMALIIECKEASQYTAIKLINGNEYQFRSAVNFVGRPLDSDPVVAQIQYTVDPAP 21658
QY 15 ----- 14
Db 21659 GIPEPSNITGNSITLWMAPESDGSEIQYTLERREKSTWVAVISKRPISETRFRVY 21718
QY 15 ---INP----- 17
Db 21719 GLTEGNEYEFHMAAGVGPASISRLIKREBVNPGPPTVVKVYDTSKTVSLEMS 21778
QY 18 ----- 17
Db 21779 KVEFDGMEIIGYITTEMCKTDLGDMHVAEACVTRVTVDLQAGEYKFRVASINAG 21838
QY 18 ----- 17
Db 21839 KGDSECVTGIKAVDRLAPELDIDANFKQTHVRAAGASIRLETAQGRPIPTAVWSKPD 21898
QY 18 ----- 17
Db 21899 SNLSLRADITHTDPSHTLVENCNRNDAGKYTLFVENNSGSKSTFVVKLDTPGPPI 21958
QY 18 ----- 17
Db 21959 TRKDVTRGSATILMDAPILDDGARLIHHYVVEREASRSRMOVISCKTRQIFKVDLAEG 22018

QY 18 ----- 17
Db 22019 VPIYFRSAVNEYGVGEPEYEMPEPIVATEQAPRRRLVDVTSKSSAVLMLKPDHDGS 22078
QY 18 ----- 17
Db 22079 RITGYLLEMRONGSDLWAEAGHTKQITTEVERIVEKTEFEFRVAKNDAGYSEPREAFSS 22138
QY 18 ----- 17
Db 22139 VIKPEQIEPTADLTGTTNOLITCKAGSPFTIDVPISGRAPKVTWKLMEMLKETDRVS 22198
QY 18 ----- 17
Db 22199 ITTKDRTTLVQDSMRGDSGRYFLLTENTAGVKTFSVTVVYIGRPGPYGPIEVSSVA 22258
QY 18 ----- 17
Db 22259 ESCVLWGEPRDGGTEITNYIVERREGTTAMOLVNSVAKTQIKYTHLKMYEGRV 22318
QY 18 ----- 17
Db 22319 SSENFGVSKLESNAPITAEHPFVPSAPTRPEYTHVSANAMSIRMEBPHDGSKITIGY 22378
QY 18 ----- 17
Db 22379 WVEKKERNITLWKENKVPCLCENYKVTGLVEGLEQFRYALNAGVSKASEASRPIMA 22438
QY 18 ----- 17
Db 22439 QNPVDAPGRPEVTVTSTVSLWSAPAYDGSKVGVYIIEKRPVSEVGDGRLKCNVTI 22498
QY 18 ----- 17
Db 22499 VSDNFTVLTALSEGDTYEFERLAKNAGVISKSESESTGPVTCRDEYAPPAELDARLHGD 22558
QY 18 ----- 17
Db 22559 LVTRAGSDLVLAAGVGKPEPKIIMTKGDKELDCEKVSQYTGKRAVAVIKFCRDS 22618
QY 18 ----- 17
Db 22619 GKTYLTVKNASGTAAVSVMKAVLDSPPCGKLVSRVTOEKCGLAWSLPQEDGAEITHY 22678
QY 18 ----- 17
Db 22679 IVERRETSRLMWVIVEGECPTLSVYVTRLLKNNEXIFRVAVANKYGGVYESEPIYARN 22738
QY 18 ----- 17
Db 22739 SFTIPSPPGIPEEVGTGKEHIIQWTRPESDGNESNLYLDRKESLRWTRVNDYVY 22798
QY 18 ----- 17
Db 22799 YDTRLKATLSMEGCDYQFRYTAANAAGNSPERSNFIISCRBPSYTPGPSPARVVDTK 22858
QY 18 ----- 17
Db 22859 HSIASLAKPMYDGDGTIVGVLEMQEKDIDQWRYVHTNATINTEETVPDLKMGKYSF 22918
QY 18 ----- 17
Db 22919 RVAAVNKNKSESESTAEIPEVERIEIPDLLEADLKTVTIRAGASRLMWVSGRPP 22978
QY 18 ----- 17
Db 22979 PVTWKSQIGDIASRALIIDTESYSLIYDKVNRVYDAGKTTIAENOSGKSKATVUVKY 23038
QY 18 ----- 17
Db 23039 DTGPGCPVAVKEVSRDSVTITWEIPTIDGAPINNYIVEKREAAARAFKTVTTKSKTL 23098
QY 18 ----- 17
```

```

Db 23099 YRISGLVBGTMHYFRVLBNINIGIGBCEPISDAVLVSEVPLVPAKLEVVDTKSTVTILAW 23158
QY 18 ----- 17
Db 23159 EKPLVDGSGHILGYVLEACKAGTERMKVVTLPVLBEHTVTLNMGEOYLFRIRAOBCK 23218
QY 18 ----- 17
Db 23219 GVSEPREVTAVTQDLRLVPLTIDLTMPQKTIHVPAGRPVELVPLIAGRPPAASWFFA 23278
QY 18 ----- 17
Db 23279 GSKLRSESRVTEVHTKAKLFIRETTRIDYGEYLELKNVGTSETIKVILIDKPGPP 23338
QY 18 ----- 17
Db 23339 TGPRIKIDELDATSITISWEPPELDGAPLSGYVVEQDAHRPQWLPVSESVTKSTFKFTR 23398
QY 18 ----- 17
Db 23399 LTEGNEVFRVNAATNRPGLSGLQSEVIEGRSSIRIPGPELQIPDVSQDQNTLTWYPP 23458
QY 18 ----- 17
Db 23459 EDDGSGQVTGYIVERKERVADRWVRVNVKVPVMTIRYSTGLTEGLEHHRVTAIINARGSG 23518
QY 18 ----- 17
Db 23519 KPSRPSKPIVAMPDIAPPGKQNPVTDPTTRISVSLAMSVPEDEGSKVTGYLIEMQKVD 23578
QY 18 ----- 17
Db 23579 QHEMTKONTPTKIRREYTLHLPOGAEYRVRVACNAGRGEPAEVPGVYKTEMLEYDP 23638
QY 18 ----- 17
Db 23639 YELDERYQBGIFVRQGVIRLTIPIKRPPICKWTEGQDISRAMIATSEHTHELVIK 23698
QY 18 ----- 17
Db 23699 EADRGDSGTDLVLEKCGKAKAYIKVAVIGSPNSPBGPLEYDIDQVRSVSWRRPADD 23758
QY 18 ----- 22
      -ANYXX-
      |||
Db 23759 GGADILGLYLERREVPKAAWYTTIDSRVGTSLVYKGLKENVEYHFRVSAENQFISKPLK 23818
QY 23 ----- 22
Db 23819 SEEPVTPKTPLNPPPPNPPEVLDTKSSVLSWSRPKDDGSRVYTYIERKETSTDK 23878
QY 23 ----- 22
Db 23879 VVRHNKQIITMTVYGLVPADEYQFRIIAQNDVGLSETSPASEPVYCKDPDKPSQBP 23938
QY 23 ----- 22
Db 23939 ELEILISKDSVTLQWKEPCEGDKKEILGYWVEYROSQDSAMKSNKERIKQFTIGL 23998
QY 23 ----- 22
Db 23999 LEATEYFRVFAENETGLSRPRTAMSIKTKLISGEAPGIRKEMKDVTYTLGSAQLISQ 24058
QY 23 ----- 22
Db 24059 IYGRPLDIDIKMYRCKELIQSRKYMSSDGRTHTLVWTEEOBDEGYTCIATNEVEGE 24118
QY 23 ----- 22
Db 24119 TSSKILLQATPOFHPGYPLKEKYGAVGSLRLHMTIGRVPAMTFHGOILLQNSENT 24178
QY 23 ----- 22

```

```

Db 24179 TIENTEHTHLVMKNVQRKTHAGKYKVOLSNVEGTVDAILDVEIODKPKDPTGPVIEAL 24238
QY 23 ----- 22
Db 24239 LKNSAVISWKPADDGSGWITINVEYKECAKEGEMOLVSALSIVTTCRIVLNENAGY 24298
QY 23 ----- 22
Db 24299 FRVSAQNTFGISDPLEYSVYIISPFEPKGPAPGPTITAVTKDSVCVAMKPPASDGAK 24358
QY 23 ----- 22
Db 24359 IRNYYLEKREKONKMWISVTTIEIRETVFSVKNLIEGLEVEFRVYKCNLGESEWSEISE 24418
QY 23 ----- 22
Db 24419 PITRSDVPIQAPHFKEHLRLNRYOSNATLVCKYGHKPIYKWRQKEIILADGLK 24478
QY 23 ----- 22
Db 24479 RIQEFKGYHOLITASVTDDAFYOVRAVNOGGSVGSFASLEVEPAKIHLPKTEGMG 24538
QY 23 ----- 22
Db 24539 AVHALRGEVVSIIKIPFSGKPPVITWQODLIDNNGHYOIVYRSFTSLVFPNGVERKD 24598
QY 23 ----- 25
      -RGI-
      |||
Db 24599 AGFYVCAKNRFGIDQKVELDVADVPDPPRGVAVSDASRVSVALTWTERASDGSKITN 24658
QY 26 ----- 25
Db 24659 YIVEKATTAERWLRVQARETRYVINLFGKTSYOFVIAENKFGSKPSEPSEPTTK 24718
QY 26 ----- 25
Db 24719 EDKTRAMNYDEBVDETREVSMTKASHSTKELYKMAIEDLGRGEGIVHRCVETSCK 24778
QY 26 ----- 25
Db 24779 TYMAKFAVKGDQVLYRKEISILINARHNLHLHESFESMEELVMIFFISGLDIFER 24838
QY 26 ----- 25
Db 24839 INTSAFEINEREIYSYHQCBAQLFLSHNIGHDIRPENIITQTRRSSTIKIIEFGQA 24898
QY 26 ----- 25
Db 24899 ROLKRGDNFRLLFTAPRYAYABVHOHDVVSATDQMSLGLVYVLLSGINPFLAETNOOI 24958
QY 26 ----- 25
Db 24959 IENINNAEYTFDEAFAKEISIEAMDFVRLLVKERSHMTASEALOHPWLKOKIERVSTK 25018
QY 26 ----- 25
Db 25019 VIRTLEHRRYHTLIKLDLNNVYSAARISCGAITSQGVSAKYKVASIEIGVSGQIM 25078
QY 26 ----- 25
Db 25079 HAVGEGGHVYVCKIENYDQSTQVWYEGVROLENSEKYEITTEDVAILYVNDITRKL 25138
QY 26 ----- 25
Db 25139 DGYRCKVYNDYGEDSYAEFLVKGVRVYDYCHRTMKKIKRRTDMLRLEPPEFTLP 25198
QY 26 ----- 25
Db 25199 LYNKTAIYGENVRFGVTTIVHPEPHVYTKSGQIKIPGDNDKKTTFESDKGLYQTLINSV 25258
QY 26 ----- 25
Db 25259 TTDDAEYTVARNNKYGEDSCAKLVTLLHPPPTDSTLRPMFKRLLANAEQOGQSVCFE 25318

```

```

QY 26 ----- 25
Db 25319 IRVSGIPPTLKWEDGQPLSGPNIIEHGLDYALHIDLPEDTGYRVATATAG 25378
QY 26 ----- 25
Db 25379 STSQALQVRLKRYKOEKSEKHEHRYOKQIDKLRMAELISTGESVPLQVAKAL 25438
QY 26 ----- 25
Db 25439 REAALYKPAVSTKVGEFRLIEEKEERKLMYPDVEPRKYKOTIEEDQRIKQFY 25498
QY 26 ----- 25
Db 25499 PMSDMWKYKIRQOYEMPGKLDVYOKRPRIRLSRWEQFYVPLPRITDQYRPRKRIK 25558
QY 26 ----- 25
Db 25559 LSQDDLEIVPARRRTPSPDYFYRRRRSLGDISDELLFLIDYLANKRTKEERLRL 25618
QY 26 ----- 25
Db 25619 EEELELGFSPSPSPHPELSLRYSPOAHVYETRNKFRYSYHIPTKASTSY 25678
QY 26 ----- 25
Db 25679 AELREHQAAYRQPKORIMAREDELLRPVTTTOLSEYKSELDFMSEKESRKKS 25738
QY 26 ----- 28
Db 25739 RQREVTEIEIEEYKISKHARESSSASRLRRRSLSPTIYELMRVSELINSRPO 25798
QY 29 ----- 28
Db 25799 PAEEYEDTERSPTEPERTRSPSPVSSERSLSREFSARFDIFGRYSMAKALTKT 25858
QY 29 ----- 28
Db 25859 SEKREYLSQOPTLHAPITTLRMKSHRYPCGONTFLINQSKPTAEKWHNGVELQ 25918
QY 29 ----- 28
Db 25919 ESSKIHNTNSVLTLEILDCHTDGTYRAVCTNKGASDYATLDYGGDYTTYSOR 25978
QY 29 ----- 28
Db 25979 RDEEVRSVFPELLTTEAVAVSPFKKTSMEWASSVREYKSOQTTRESLSYEHSASAE 26038
QY 29 ----- 28
Db 26039 MNSALEEKSLEKSTTRIKITTLARILITKPRSMTVYEGESARRSCDTGEPVPTVWL 26098
QY 29 ----- 28
Db 26099 RKGQVLSARHQVTTTKYKSTFEISSVQASDEGNYSVVENSEGQAEFTLTIQKARY 26158
QY 29 ----- 28
Db 26159 TEKAVTSPPRVKSPEPRVKSPEAVKSPKRVKSPESHPRVAPTETKPTPREKQHLVPS 26218
QY 29 ----- 28
Db 26219 APPKTIQFLAKASKELIAKLVGVESVLRKAEVYTWYKDGKRLKENGHPHYHSDGYE 26278
QY 29 ----- 28
Db 26279 LKINMLTESDQGYVCEISGEGSTKTNLOFMGAQFKSIHEKYSKISETKKSQKTTST 26338
QY 29 ----- 28
Db 26339 VTRKTEPRKPEPISSKPVIVTGLDQTVSSDVANFAVKANGEPPTAIWTKDGRAITOG 26398

```

```

QY 29 ----- 28
Db 26399 GKXKLSDEKGGFFLEIHKTDSDSGLYCTVKNASGVSSSCKLTIKAKTQAKVSTQ 26458
QY 29 ----- 28
Db 26459 KTSEITPQKAVQEEISQKALRESEIKSEAKSQEKLAKREASKVLISEEVKSAATS 26518
QY 29 ----- 28
Db 26519 LEKSIYHEITTSQASEVRYTHAETKASTQMSINEGORLVKANINAGARDVKNVNGV 26578
QY 29 ----- 28
Db 26579 ELTNSEYRYGVSGDQTLTIKQASHRDEGLITCISTKEGIVKQYDLTISKELSDAPA 26638
QY 29 ----- 28
Db 26639 FTSQPRSONINGQVLFCTCEISGEPSEIEMFKNNLPISISSNYSISRSRYVSLERIN 26698
QY 29 ----- 28
Db 26699 ASVSDSGKTTIAKKNFRCQCSATASLAMPVLYEPPREVLRTSGDLSGSSQSYQM 26758
QY 29 ----- 28
Db 26759 SASQEPASFSFSSSSASMTKFAKMSAQSMSSMOESFYEMSSSPMGISNMTQLESS 26818
QY 29 ----- 29
Db 26819 TSKMLKAGIRIPRILEALPDISIDEKVLTVACAFGEPTPEYTWSCGGRKIHSDQOG 26878
QY 30 RF 31
Db 26879 RF 26880

```

RESULT 12

Q8WZB3

AC Q8WZB3 PRELIMINARY; PRT: 26926 AA.

DT 01-MAR-2002 (TREMBLrel. 20. Created)

DT 01-MAR-2002 (TREMBLrel. 20. Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21. Last annotation update)

DE N2B-tilin isoform.

GN TJN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20309627; PubMed=10850961;

RA Freiburg A., Trombitas K., Hell W., Gazorla O., Fougereuse F.,

RA Centner T., Kolmerer B., Wilt C., Beckmann J.S., Gregorio C.C.,

RA Granzler H., Labelt S.;

RT "Series of exon-skipping events in the elastic spring region of titin

RT as the structural basis for myofibrillar elastic diversity.";

RL Circ. Res. 86:1114-1121(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21573839; PubMed=11717165;

RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,

RA McNabb M., Wilt C.C., Labelt D., Gregorio C.C., Granzler H.,

RA Labelt S.;

RT "The complete gene sequence of titin, expression of an unusual -700

RT kDa titin isoform and its interaction with obscurin identify a novel

RT Z-line to I-band linking system.";

RL Circ. Res. 89:1065-1072(2001).

DR EMBL; AJ277892; CAD12455.1; -

DR InterPro; IPR000283; Cytok_receptor_2.

DR InterPro; IPR000719; Euk_kinase.

DR InterPro; IPR00577; FGGT_kin.

QY 6 ----- 5
Db 4199 SDEKQDESLKPESEKESSESSETEEVATVIOEAGGFIKEDQPMHITPLVDTVSEGD 4258
QY 6 ----- 5
Db 4259 IVHLTTSITNAKEVNTVEFNKLVPSDEKCLODQNTYTLVDKNTEDHQGEVCEALN 4318
QY 6 ----- 5
Db 4319 DSGKTATSAKLTVYKRAPIYKRIEPLVALGHLAKFTCEIOSAPNRVROWKAGREIY 4378
QY 6 ----- 5
Db 4379 ESDKCSIRSSKYSISLEILRTQVVDGCEYCKASNEYSVSTATILTVVGGKKVRL 4438
QY 6 ----- 5
Db 4439 LPERKPEKEEVVLKSVLRKRPBEBEPKVPKLEKVKKPAVEPPPPKPEVEEVPYV 4498
QY 6 ----- 5
Db 4499 KREKKIPEPTKVPKIPALPLPAPBEPKPEAEVKTIKPPVPEPPTIAPVTVVVGK 4558
QY 6 ----- 5
Db 4559 KAKAKAPKEEAAKPKGPIKNGPKTPSPITAEERRKLRPGSGEKKPPDEAPFYQLKAVPL 4618
QY 6 ----- 5
Db 4619 KVKKEIKDILITSESEVSSAIECLVSPSTAITTMMKGSNIRESPKHREFIADGKORL 4678
QY 6 ----- 5
Db 4679 HIIDVLDAGEYTCVLRGKNEKSTAKLVVEELPVRFVKTLEEVTVVKGQPLVLSCE 4738
QY 6 ----- 5
Db 4739 LNKERDVWRKDGKIYKVRGRIVPGVIGLMRALTTNDADDTAGTYTVVENANMLECS 4798
QY 6 ----- 5
Db 4799 SCVYVEVLRDMVLKPIRQHVAKPGTAFACDIANDJTNIKFKGYDELPAEPNDKTEI 4858
QY 6 ----- 5
Db 4859 LRDNHLYLIKINAMEPIAEYAVEIEGRYPAKLTIGEREVELLPIDVITYEKESAS 4918
QY 6 ----- 5
Db 4919 FDAISEADIPGOWKLKGLLRPSPTCEIKABGKRFLLTRVKLDOAGEVLVYQALNAT 4978
QY 6 ----- 5
Db 4979 TAILTVKEIELDPAVPLKDVTVERRQARFECVLTPREANVWSKGPDIKSSDKFDITAD 5038
QY 6 ----- 5
Db 5039 GKXHLIVINDSOPDDEGVYTAIEGAKTISARLFVTGIRLKFMSPLEDQTVKEGETATFVC 5098
QY 6 ----- 5
Db 5099 ELSHEKMHVWVRKNDAKLHSTRVLISSEGTKHKLBMKEVTLDDISQIKAOVKELSTAQ 5158
QY 6 ----- 5
Db 5159 LKVLADPYFTVKLHDKTAVERDEILKCEVSKDVVKWFKDGEELVSPKYSIKADGIR 5218
QY 6 ----- 5
Db 5219 RLKIKKADLKDGKGYCDGCTDKRANTYBARLITKVKPLYGVEVGETAHEFEIELS 5278

QY 6 ----- 5
Db 5279 EPDVHGWMLKQGPLTASPCEIIEGKKHILILNCOIGMTGEVSPQANAKSANTKV 5338
QY 6 ----- 5
Db 5339 KELPLIFITPLSDVKVEKDEAKFCEVSRBPKTPRMLKGTQELTGDDRFLINKDTKHS 5398
QY 6 ----- 5
Db 5399 MYKSAFEDEAKYFEADKHTSGKLIIEGIRLFLPLDVTAKESAVFVELSHD 5458
QY 6 ----- 5
Db 5459 NIKVWFKNDRLHTTRSVSMODEKTHSITPKDLSIDTSQIRVEANGMSAKLYLE 5518
QY 6 ----- 5
Db 5519 GDPYFTGLQDYTGVEKDEVIIQCEISKADAPVKWFKDGEIKPSKNAVIAKDGKKMLI 5578
QY 6 ----- 5
Db 5579 LKALKSDIGQYTCDCGTDKTSGLDIEDREIKIVPLHSVEVMEETARETEISEDDI 5638
QY 6 ----- 5
Db 5639 HANWMLKGEALLQTPDCEIKEGKTHSLVHNCRLDQGTGVDFQANVKSANHLRVKPRV 5698
QY 6 ----- 5
Db 5699 IGLRLPLDVTAGETATFCELSYEDI PVEMYLKSKLEPSDKVVRSGKVHTLLR 5758
QY 6 ----- 5
Db 5759 DVKLEDAGEVOLTAQDFTHANLTVKEPPVEFTYPLEDQVEGATAVLCEVSRENKV 5818
QY 6 ----- 5
Db 5819 KWFKNGTIELSKKYEIVADGRVRLVHDCPTBEDIKTYTDANDFKTSCNLNVPVHVE 5878
QY 6 ----- 5
Db 5879 FLRPLTDQVREKEMARECELSRENKVKWFKDGAIBKGGKYDIISKGAVRILVINKC 5938
QY 6 ----- 5
Db 5939 LLDDEAEXCEVRNARTSGMLTVLEEAVFYKNLANIEVSETDIKLVCYSKPGAEVIM 5998
QY 6 ----- 5
Db 5999 YKGEELIETGRVEILTEGRRRIIVTONAHLEDAGNVCNRLPSSRTDGKVVHLEAEFI 6058
QY 6 ----- 5
Db 6059 SKPONLELBEKAEFVCSISKESFPVQWKRDLTLESGBKTYIADGKKRVLVKDATL 6118
QY 6 ----- 5
Db 6119 QDMGTYYVNGAARAANHLVIEKLRIVPLKTRVKEODEVYVNCENVTEGAKAKFRN 6178
QY 6 ----- 5
Db 6179 BEAIFDSSKYIILQKDLVYTLIRIDAHLDQANVNSLTNHRGENVKSANLIVEEDLR 6238
QY 6 ----- 5
Db 6239 IVEPLKOLETMEKKSVPFWCKVNRNLANTLAKWTKNGEEVPPDNRYSVRVDRYKXMLTKDC 6298
QY 6 ----- 5
Db 6299 GFPDEGEYIVTAGOKSVAELLIEAPTEFVEHLEDQTVTEFDDAVESCOLSREKANVM 6358
QY 6 ----- 5

Db 6359 YRNGREIKBEKKYKFEKDSIHRLIIKDCRDECEYACGVEDRKRARLVEIPEI 6418
QY 6 ----- 5
Db 6419 RPPODILEAGADVFLAELNKDKVEYOMLRNMVVVGGDKHQMSEKIHRLQICDIPR 6478
QY 6 ----- 5
Db 6479 RDGGERFIADKEARAKELAAAPKIKTADODLVVDGKPLTMVVPYDAIYPAEAEMER 6538
QY 6 ----- 5
Db 6539 ENELSTKITDTAEQTSFIFLEKKGDKGRYKIVLQNKHKAEGFINLKVIDVPGVRN 6598
QY 6 ----- 5
Db 6599 LEVTEFDGEVSLAWEPLTDGSKIIIGVYVERBRDIKRTIWLATDRAESCEPTVGLQK 6658
QY 6 ----- 5
Db 6659 GGEYELFRVSARNRVGTGEVETDNPVEARSKTDVGPPLNTITDVNRCVSLTWEPR 6718
QY 6 ----- 5
Db 6719 YDGAETNTVYIELDKTSIRMDIAMTVRAEDLSAFTVDYVEGQESFRVAQNRIGVGR 6778
QY 6 ----- 5
Db 6779 PSAATPEVAVADPIERSPPVNLTSDQTSVOLKWEPLKDGSPILGIYIERCEBK 6838
QY 6 ----- 5
Db 6839 DNWIRCMKLVPELTYSVTGLEKGNKLYRVSANENKAGVSDPSILGLTADAFVEPTM 6898
QY 6 ----- 5
Db 6899 DLSAFKDGLEIVPNPITIIIVPSTGYPRPATWCFGDKVLETDGRVAKMTLSAYAEVLIS 6958
QY 6 ----- 5
Db 6959 PSESDKGIYTLKLENNKVIISGEIDVNVJARPSAPKELFGDITTKDSVHLTWEPPDDG 7018
QY 6 ----- 5
Db 7019 GSPITGYVEKEVSRKWTVMQVTDLEFTVPLVQGEYLFKVCARMKCGPEBAYV 7078
QY 6 ----- 5
Db 7079 DEPVNMSTPATVPDPENVKWRDRTANSIFLTWDPKNDGSGRIKGYIVERCPGSDKW 7138
QY 6 ----- 5
Db 7139 ACGEPAETKMEVTGLEGGKWTATRVKALNRQASKSPRPEIEQAVDQEAPEIIFLDV 7198
QY 6 ----- 5
Db 7199 LLAGITVAGTIELPATVTGKPEPKITWTAKADMILKODKRITIEVNPCKSTVIYVSKR 7258
QY 6 ----- 5
Db 7259 SDTGTYIIEAENVCGRAIYAVENVLDKPGPAAFDITDVNESCCLTWNPPRDDGSKI 7318
QY 6 ----- 5
Db 7319 TNYVERATDSEVWHKLSIVKDTNFKATKILIPNKEIIFVAAENMTGVGEIPAQAPIT 7378
QY 6 ----- 5
Db 7379 AKYQDPGPPRLPEPDTIDAVTLFWCEPDDGSPITGYWERLDPDTDKWVRCNM 7438
QY 6 ----- 5

Db 7439 PVKDTYHVKGLTNKKYRFRVLAENLAGPGRPSKSTEPILINDIDPPWPPGKPTVKDV 7498
QY 6 ----- 5
Db 7499 GRTSVRLMWTKEPHDGAKIESYIEMLKITGTDENVRAESVPTTOHLLPELMEGQESY 7558
QY 6 ----- 5
Db 7559 RVRANVKAESBSPSPDVLCREKLYPPSPRMLEVINITKNTADLKMTVEKDGSP 7618
QY 6 ----- 5
Db 7619 TNYIVEKDVRRKGMQVDTYVTKDCTVETPLTBSLTVFRVAENALIGSDYTEIEDSV 7678
QY 6 ----- 5
Db 7679 LAKDTFTTGPFPYALAVDVTKRAVDLKWEPPKNDGPRIORVYIEKKERLGTWVAKG 7738
QY 6 ----- 5
Db 7739 TAGPCNFRVTDVIGTEVOVRAENAGVHPSEPTIELSDPTSPFPDILHVT 7798
QY 6 ----- 5
Db 7799 AGRKHIALAMKPEKNGSPILIGYHVMCPVTEKMRVNSRPIKDLKFVEGVPDKE 7858
QY 6 ----- 5
Db 7859 YVLRVAVNAIGVSEPSISENVVAKDPCKPTIDLETHDIIVIEKSLIPVFPBAPV 7918
QY 6 ----- 5
Db 7919 PTVSMHKGKEYKASDRILMKNDHISAHLEVPKSVRADAGIYTTLENKLSATASINVK 7978
QY 6 ----- 5
Db 7979 VIGLPECKDIASDITTSCKLTWEPPFDGTPILHVLERRAGRTIIPVMSGENK 8038
QY 6 ----- 5
Db 8039 LSWTVDLIPNGEYFPRVAVNKVGGEYIELKNPVAQDPKPPDPVEVHNPTAE 8098
QY 6 ----- 5
Db 8099 MTITWKPPLYDGSKIMGYIIEKIAKGERMRKCNHLPILTYAKLEGGKEYQFVR 8158
QY 6 ----- 5
Db 8159 AENAGISEPSRATPPTKAVDPIDAPKYLRTSLVKRGDEIALDASISGSPYTIWIK 8218
QY 6 ----- 5
Db 8219 DENVIYEEIKRAAPLYRRRGEVQEEPEVLPLOTSLDINSKKSQULVRDSLPRD 8278
QY 6 ----- 5
Db 8279 HGLYMKVENDGIAPCTSVLDTGPPINFEEDIRKTSVLCCKWEPPLDDGSEIIN 8338
QY 6 ----- 5
Db 8339 YLLEKDKTRPDSMTIYVSTLRHCKYSVTKLIEGKEYLFRVARNRPGPPCVSKPLV 8398
QY 6 ----- 5
Db 8399 AKDPGPPADPKPIVEDYTSNMLYKNNEPKDNGSPILGYWLEREVNTHSMHVNSL 8458
QY 6 ----- 5
Db 8459 LNALKANVDGLLEGLTYVERVCAENNAGPKFSPSDPKTAHDPIISPFGPIPRVYDTSS 8518
QY 6 ----- 5
Db 8519 TTIELEMPAPFNGGEIYGVFDKQLVGTNEMSRCTEKIMIVROYTYVEIREGADYKLR 8578

QY	6	-----	5
Db	8579	VSANMAGEGPGHPTQPVTAEPQEPAPVELDVSKGIQIMAGKTLIRIPAVTGRPVPT	8638
QY	6	-----	5
Db	8639	KVMTKEGELDKDRVVDNVTGKSELIIKDALRKDRGVITATNCGSKFAARVEVFD	8698
QY	6	-----	5
Db	8699	VPGVLDLKPVTYNNKCKLNMSPDEDDGSEITGFIIEKKDAKMTWROPITENSKCD	8758
QY	6	-----	5
Db	8759	ITGLLEQGEYKFRVIAKNGCGPVEIGPILAVDPLGPPPTSPERLTYTERKSTITLDM	8818
QY	6	-----	5
Db	8819	KEPNSNGSPIQGIIEKRRHDKPDERVNRKLCPTTSELVENLDEHOMTEFFVKAVNET	8878
QY	6	-----	5
Db	8879	GESEPSLPLNVVIQDDEVPTIKLRISVRGDTIKVAKGEVHIIPADVTGLPMKIKESKN	8938
QY	6	-----	5
Db	8939	EYVIEKPTDALQITKEEVSRSEAKTELSTPKAVREDKGYTVTASNRLGSVFNNHVEYV	8998
QY	6	-----	5
Db	8999	DRSPRNLAVTDIKAESCLYTMDAPLNDNGSEITHYVIDKRDASRKAEMEEVNTAVE	9058
QY	6	-----	5
Db	9059	KRYGIMKLIPNGOYEFERAVRANKYISDECKSDKVVIQDPERLPGPPKPKVLARTGSM	9118
QY	6	-----	5
Db	9119	LVSMTPLDNGGSPITGYMLEKREBSPIWRSRAPITRVGLKGEVNPRLLEGVXYQ	9178
QY	6	-----	5
Db	9179	FRAMAINAGIGPSESPDEVAGDPLFPGPSPCEVMDKTKSSISLGMKPPAKDGGSP	9238
QY	6	-----	5
Db	9239	IKGYIVEMOEEGTTDMKRVNEDPKLITTCQCVVNLKELRYFRVKAVERNAGESEPSDT	9298
QY	6	-----	5
Db	9299	TGEIPATDIOEPEVEFIDIGADCLYCKAGSQIRIPAVIKGRPTPKSSWEFDGKAKKAMK	9358
QY	6	-----	5
Db	9359	DGVHIDPDAQLETAENSVIIIECKRSHTGKYSITAKNKAGQKTANCVRKVMVDPGPP	9418
QY	6	-----	5
Db	9419	KDLKVSADITRGSCRLSKMPPDDGDRIGYIEKRTIDGKAMTKVNPDCGSTTEVPDL	9478
QY	6	-----	5
Db	9479	LSEQJFFRVRARENRFGIGPVETIQRTARDPYPPDPPIKLKIGLITKNTVHLSMKRP	9538
QY	6	-----	5
Db	9539	KNDGSPVTHYIYVECLAMPDTGTKEAMRQCNKRDVELOQTVEDLVEGGEYFRKAVN	9598
QY	6	-----	5
Db	9599	AAGVSKPASTVGPDCQGRDMPSPSIDLKEFMEVEBGTNNYAKIKGVFPPTLTWFKAPP	9658

QY	6	-----	5
Db	9659	KRPDNKEPVLVDTHVKNLVVDCTILVIPQSRSDTGLYTTAANNLGTASKEMRLNVIG	9718
QY	6	-----	5
Db	9719	RPGPVPIKESVSADQMTLSMFPKDDGSKITNYIEKREANRKTWVHSSPKECT	9778
QY	6	-----	5
Db	9779	YTIPLLEGEHYVRIMAONKYGIGEPIDSEPETARNLFSVGABDKPTVSSVTNSMTV	9838
QY	6	-----	5
Db	9839	NMEEPYDGSPTVGTWLEMKDTTSKRKRVRNRPDIKANTLGVSYKVTGLLEGSDYQFRV	9898
QY	6	-----	5
Db	9899	YAINAGVPASLPSDPATARDPIAPGPPFPKVTMDTKSSADLEMSPLKDGGSKVTGY	9958
QY	6	-----	5
Db	9959	IYIEKEGKEEWEKGDKEVNGTKLVTTGLKGADEYKFRVSAVINAGIGEPEVTVIEM	10018
QY	6	-----	5
Db	10019	KDRLVSPDLQLDASVRDRIVHAGVIRILIAVSGKPPPTVWNNERTLPQEAETIETA	10078
QY	6	-----	5
Db	10079	ISSMVIKNCQSHOGVYSLAKNAGEKRTUIVIDVDPGVPGTPELANLINESCL	10138
QY	6	-----	5
Db	10139	TWFSPEDDGSPITNYIEKRESRDRANTPYTYTTRONATVOGLIOGKAVFRIAENS	10198
QY	6	-----	5
Db	10199	ICMGFVETSEALVIREPITVPERPELEVKENTVTLTMNPKYDGGSEIINYLES	10258
QY	6	-----	5
Db	10259	RLIGTEFHKTNDNLRSKTYVGLKEGDYEVKRSANIVGQKPSFCTKPYTCDEL	10318
QY	6	-----	5
Db	10319	APPTLHDFRDLIRVGEAFALIGRYSKPKPKVSMKEADADLEDDRTIKITPATLA	10378
QY	6	-----	5
Db	10379	LEKIRAKRSDGKCYCVVENSTGSRKFCQYVNVDRPGPVGVSFDEVTEDYVISMKP	10438
QY	6	-----	5
Db	10439	PLDDGSKITNYIEKKEVGDMMPVTSASAKITCKVSKLLEBKDYIFRIHAENLYGIS	10498
QY	6	-----	5
Db	10499	DPLVDSMKANDRFRVPADPQPIYEVTKDSALVTWNNKPHDGKPTINYILLEKRETMK	10558
QY	6	-----	5
Db	10559	RMARVTKDPIHPYTKFRVBDLLEGCQYEFVSAENEGIDGSPSPKVPYAKDPIAKPSP	10618
QY	6	-----	5
Db	10619	PVNPEADITGNSVDLTWQPPRHGSGKILGYIYEQVGDHEMRANRHPESCPETIKYK	10678
QY	6	-----	5
Db	10679	VTGLDQGTYFRVLAVNAGESDPAHVEPEVLVKDRLEPPELLIDANMARQEHIVGDT	10738
QY	6	-----	5

Db 10739 LRLSAIIKGVFPFKVTKKKEDRDAPTAKRIDVTPVSGSKLEIRNAHEDGYSLVJENDA 10798
QY 6 ----- 5
Db 10799 GSKTVSVKVLVLDKPGPBDLEVESEIRKDSCLTWKREPLDDGGSVITNTYVERRDVSAQ 10858
QY 6 ----- 5
Db 10859 WSPLSATSKKSHFAKHLNCGNOLFRVAENOGKGFVEYETPKIKALDPLHPGPPKD 10918
QY 6 ----- 5
Db 10919 LHHVDVKTESLVWNNKPRDGGSPITGVLYEOEGTODMIKFKTVNLECVVTLQOG 10978
QY 6 ----- 5
Db 10979 KTYRFRVKAENIVGLPDTTIPLECOEKLVPSPVELDKLIEGLVYKAGTVRPAIIR 11038
QY 6 ----- 5
Db 11039 GVPVPTAKWTIDGSEITDEHYETDNFSSVLTIKNCLEDRDGEYOITVSNAAGSKTVA 11098
QY 6 ----- 5
Db 11099 VHLTVLDPGPPPTPINILDPTEPHMTISWQPKDDGSPVINYIEKODTRKDTMGVVS 11158
QY 6 ----- 5
Db 11159 SGSSRTKLKIPHLQKGEYFVRVAENKIGVPPLDSTPTVANKHSPSPSGKPVYTDI 11218
QY 6 ----- 5
Db 11219 TENAATVSWTLPKKDGSPITGYMERREVTGKWRVNTKPIADLKFRVTLGVEGNTEF 11278
QY 6 ----- 5
Db 11279 RVFAENLAGLSKSPSSDPKACRPKIPGPPINPKLKDSREYADLVWTKPLSDGSP 11338
QY 6 ----- 5
Db 11339 LGYVVECOKPGTAOMNRINKDELIRQCAFVYPCGLENETRFRKANAIVGEGEPRELA 11398
QY 6 ----- 5
Db 11399 SVIAKDLHPPEVELDYTCRDVITVRVQOTIRILARVGRPEPDITWTEGKVLVREKRV 11458
QY 6 ----- 5
Db 11459 DLIODLPRVELQIKAVRADHKYIIISAKNSSGHAOGSAIVNLDRPQCQNLKVTNVT 11518
QY 6 ----- 5
Db 11519 ENCTISWENPLDNGSEITNTIYERKPNOKGMSIVASDTKRLIKANLANNEYFRVC 11578
QY 6 ----- 5
Db 11579 AENKVGVPITETKPIIAINPIDRPGEPENIADKGFVYLVKWRPPYDGGSPMLSY 11638
QY 6 ----- 5
Db 11639 HVERRLKGSDDMERVHKSIKETIHYMDRCVENOYIEFRVOTKNEGGSWMVTEEVYVK 11698
QY 6 ----- 5
Db 11699 EDLQKPVLDLTLGVLTKAGDTIRLEAGVGRKPFPEVAMTKDADATDLRSPVAKIDTR 11758
QY 6 ----- 5
Db 11759 ADSKFSLTJAKRSDGCKYVVTATNTAGSFAVATVNLDPKGPVRLKIVDVSSDRCTV 11818
QY 6 ----- 5

Db 11819 CWDPPEDGGEIONYLIEKCETRKMVNSTYSATVLAPEGITVTRLIEGNEXIFRVAENK 11878
QY 6 ----- 5
Db 11879 IGTGPTESKPVIAKTKDKGRDPPPEVTVKSEEMTVANNPPEYOGKSTIGYFLEK 11938
QY 6 ----- 5
Db 11939 EKHSTRWVPVNSALPERRMVONLLPDHEYQFRVAENEIGEPSLPSRPVAKDP 11998
QY 6 ----- 5
Db 11999 PPGPTNFRVVDPTTKHSITLGWKPVDGAPITIGVYEMRPKTIADSPDEGMRCAAAA 12058
QY 6 ----- 5
Db 12059 QLVKKEFTVTSIDENOYEFRCAONOVGIGRAPALKEAIKPKELLEPPPEIDLASMRKL 12118
QY 6 ----- 5
Db 12119 VIVRAGCPILFAIVRGRPAKVTWRKVGIDNVVRKGVLDVDMAPLVIENSTRDSDGK 12178
QY 6 ----- 5
Db 12179 YSLTVNPAGEKAVTVNVRVLDTPGVPVDLKVSDYTKTSCHVSNAPPENDGSOYTHIV 12238
QY 6 ----- 5
Db 12239 EKREADRTWSTVTPREVKTSFHTNLIYPGNEYFRVAVNEYGVPVTDVCPVLASDP 12298
QY 6 ----- 5
Db 12299 LSEDPBRKLEVTENKNSATLAMLPLNDGSAKIDGYITSREBEOPADRWTEYSVKD 12358
QY 6 ----- 5
Db 12359 LSLVVTGLEKGYKFRVAARNAVGSLEPREAGYTEAKEDLLPKIIMPQIITKAGK 12418
QY 6 ----- 5
Db 12419 LRLEAHVGVKPHPTCKWKKEDEEVTSSHAVHKAADSSILLIKDVTBKDGYSVLAEN 12478
QY 6 ----- 5
Db 12479 SSGTDQIKVVMADAPGPPDFDISDADACSLSMHIPLEDGSSNITNYIEKCDVS 12538
QY 6 ----- 5
Db 12539 RGDWYATLASVTKTSCRVGKILPGOETIFRVAENRFGISEPLTSPKXVAOFPEGVSEP 12598
QY 6 ----- 5
Db 12599 KNAKVTKVNCDFVAMDRDSDGSPITGYLIERKERNSLMWKANDTLVRYSTEYPCAG 12658
QY 6 ----- 5
Db 12659 LVEGLEYSFRYIALNKAGSSPPSKPEYVTAMPVDPPEKPEVIDVTKSTVSLIAPKX 12718
QY 6 ----- 5
Db 12719 DGGSKIIIGFVENACKLPGDKWVRNCNTAPHQIPOEETATGLEKAQOYOFRAIARFVANS 12778
QY 6 ----- 5
Db 12779 PPSPPDPVTILAENVPRIDLSVAMKSLIVKAGTNCIADATVEGKMPVTSNRKDGTL 12838
QY 6 ----- 5
Db 12839 LKPAEIKAMORNICLTLELFSVNRKSGDYITTAENSSGSKSANTIKLVLDKPGPPASV 12898
QY 6 ----- 5
Db 12899 KINKMTSDRAMLSWEPPLDDGSEITNTIYDKRETSRPWMAQVSAVTPITSCSVEKLEIG 12958

QY 6 ----- 5
Db 12959 HEYOFICAENKYGVDVTEPAIAKNPYDPGRCDPVIISNTKDHMTVSKMPADGG 13018
QY 6 ----- 5
Db 13019 GSPITGYLEKRETOAVNMTKVNKKPIIENTLAKTGLQEGTEYEFRTALINKAGPGKPSD 13078
QY 6 ----- 5
Db 13079 ASKAAYARDOPYPAPAPAFKYVDTRSSVSLSMGKPAVGGSPICIGLYEVKRADSDMW 13138
QY 6 ----- 5
Db 13139 VRCNLPONLQKTRFEVTLMEDTQOFRVYAVNKIGYSDPDVDPKHHPKDLIPPEGL 13198
QY 6 ----- 5
Db 13199 DADLRKTLIRAGVTMLYVPVGRPPPKITMSKPNVNLDRIGLDIKSTDFDTFLCEN 13258
QY 6 ----- 5
Db 13259 VNKYDAGKYITLLENSCGKRETIYVKVLDTPGPVNVYKELSKDSAVYTWEPPIIDGG 13318
QY 6 ----- 5
Db 13319 SPIINVVQKRDABKRSWSTVTTECSKTSFRVANLEGKSYPFRVAENEYIGDDETR 13378
QY 6 ----- 5
Db 13379 DAVKASQTPGPVVDLKVRSYSSKSSCSIGMKKPHSDGSRIGIVDPFLTEENKQWRVMS 13438
QY 6 ----- 5
Db 13439 LSLQYSAKDLTEGKEYTFRVSAENENGEPSEITVAVARDVAPDLDKGLPDLCYLAK 13498
QY 6 ----- 5
Db 13499 ENSNFKLPIKPKAPASVSKGGEPLATDTRVSVESAVNTLLIVDCQSDAGKYTI 13558
QY 6 ----- 5
Db 13559 TLKNVAGTEGISTIKYVGKPGIPTGPIKFDEYTAEMTLKMAPKDDGSEITNTILEK 13618
QY 6 ----- 5
Db 13619 ROSVNNKWTCASAVOKTTFRTVRLHEGMEYFRVSAENKYGVEGELKSEPIVANHPEVY 13678
QY 6 ----- 5
Db 13679 PDAPEPNTVDVNRHDSVSLTWTDPKKTGGSPITGHLFEKERNSLMKRANKTPIRMRDF 13738
QY 6 ----- 5
Db 13739 KYTGLTELEFERVWALINAGVGRSLPSEPVVALDPIPPGKPEVINITRNSVTLTWT 13798
QY 6 ----- 5
Db 13799 EKRVDGKHLTGITYEKRDLPSKSMKANHVNPBCAFTVTDLVGKGKYEFRIRAKNTAG 13858
QY 6 ----- 5
Db 13859 AISAPSESTETIICKDEYAPTIVDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSMSK 13918
QY 6 ----- 5
Db 13919 AKKDINPDSITQITSTPSSMLTIKYANKKDAGEYITATNPFGTVEHVKTVDVDPGP 13978
QY 6 ----- 5
Db 13979 PGPVEISNVAEKATLTWTPLLEDGSPKSYILEKRETSRLMTVSEDIQSCRHVATK 14038

QY 6 ----- 5
Db 14039 LIQGNETIFRVSAVNHGCEPVQSEPVAMNDRPGPPPEKPEVSNVTKNTATVSWERP 14098
QY 6 ----- 5
Db 14099 VDDGSEITGYHERREKSLRWRAIKTPVSDLRCKVTGLQEGSTIEFRVSAENRAGIG 14158
QY 6 ----- 5
Db 14159 PPSSEASDVLMKDAVYPGPPSPNPHVTDTTKKASLAWKPHYDGLTGTGVVEHQKVG 14218
QY 6 ----- 5
Db 14219 DEAWIKDITGTALRITQFVVPDLOTREKYNFRISAINDAGCEPAVIDVEIVEREMAPD 14278
QY 6 ----- 5
Db 14279 FELDELRTLLVVRAGLSIRIFVPIKGRPAPEVWTKNIMLNKRANIENTESFTLLIIP 14338
QY 6 ----- 5
Db 14339 EGNRYDTGKFVNTIENPAGKSGEVNVRVLDTPGPVNLRLPTDITKDSVTLHMDLPLIDG 14398
QY 6 ----- 5
Db 14399 GSRITNYIYEKREATRKSSTATTCCHKCYKYTGLSGCEYFRVMAENEYIGIPELET 14458
QY 6 ----- 5
Db 14459 TEPVKAESPSPDLSINIMDTIKTSVSLAWPKPHDGSKITGYVIEAQRKSDQWTHIT 14518
QY 6 ----- 5
Db 14519 TVKGLCEVVRNLTGEBEYTFQVMAVNSAGRSAPRESRPVYKEQOTMLPELDLRGIYQKLV 14578
QY 6 ----- 5
Db 14579 IAKAGDNIKVEIPVLRGPKPTVWTKGQDLKQORVNEFTTANSTILINIECVRSQSR 14638
QY 6 ----- 5
Db 14639 YPLTARNIVEGVDTITQVHDIPGPTGPIKDEVSDFVTFMDEPNDGVPISNYV 14698
QY 6 ----- 5
Db 14699 VEMROTSTWVELATVYITTYKATRLTGLDYQFRVKAQNRKGVGPGITISACTVANTP 14758
QY 6 ----- 5
Db 14759 FKVPGPSTPOVTAVKDSMTISWHEPLSDGSPILGYHVERKERNGILMOTYSKALVPG 14818
QY 6 ----- 5
Db 14819 NIFKSSGLTDGIAEFVRIANMAGKSPKSPSEPMALDPIDPGKVPBLNTRHTVTL 14878
QY 6 ----- 5
Db 14879 KWAPETYGFKITSYIYEKRDLPNGRMLKANFSLNLENETYSGLTEDAAEFERYIAKN 14938
QY 6 ----- 5
Db 14939 AAGAISSPSESDAITGRDVEAPKIKVDYFKDVTILKAGEAFRLBADVSGRPPTMEW 14998
QY 6 ----- 5
Db 14999 SKDGKELEGTAKLEIKIADSTNLVNDSTRDSGATTLTATNPGRAKHIFNVKYLDRP 15058
QY 6 ----- 5
Db 15059 GPPGPLAVEVTSKCVLSWFPPLDDGAKIDHYIVQKRETSRLAMTVNASVQYTKLK 15118
QY 6 ----- 5

Db 15119 VVKLLKGNEXYIRMAVANKYGVGEPLSESPVLAVNPYGPDPKPEVTITIKDSMVCW 15178
QY 6 ----- 5
Db 15179 GHPPDSGGSEIINYIVERDKAGQRMKCNKKTLDLRKYVSGLTGEGHEFFRMAENNA 15238
QY 6 ----- 5
Db 15239 GISAPSPSPFYKACDPTVERPGPNRVLDTSRSSISIAAMKPIYDGGSEITGMVEIA 15298
QY 6 ----- 5
Db 15299 LPBEDMOIVTPPAGLKATSYITITGLTENQEKIRIYAMNSGIGEPALVPGTPKAEDRM 15358
QY 6 ----- 5
Db 15359 LPPEIELDLARKVYIRACCTLRFPPIKGRPAPEVKWARDHGESLDKASIESTSSYTL 15418
QY 6 ----- 5
Db 15419 LIVGNVNRDSCXYIITVENSOGSKSAFVAVNRVLDTPPGPDLKVEYVTKTSVTLTWDP 15478
QY 6 ----- 5
Db 15479 LLDGSKIKNYIVEKRESTRKAYSTVATNCHKTSMKVDLOEGCSYFVLAENEXIGL 15538
QY 6 ----- 5
Db 15539 PAETASVKASERPLPGKITTMDVTRNSVLSWEKPEHDGSRILGYIVEMQTKGSDRW 15598
QY 6 ----- 5
Db 15599 ATGATVATATITGLIGSEYSFRSAQNEKGISDPRLQSVPIAKDLVTPAFKLLFN 15658
QY 6 ----- 5
Db 15659 TFFVLAGEDLKVDVPIGRPTPAVTHKDNVPLKQTRVNAESTENSLTITKDACREBY 15718
QY 6 ----- 5
Db 15719 GHVYVKLITNSAGAIETFLNVIIDKRGPPPGPVAKMEVTAIDSTILSNKPKYDGGSSINN 15778
QY 6 ----- 5
Db 15779 YIVEKBDTSTTQIVASATVARTTIKACRLKTCGEYQFRIAENRYGKSTYLNSEPTVAQ 15838
QY 6 ----- 5
Db 15839 YPFKVPGPPTPVVTLSSRDSMEVQWNEPISDGSRYIGYHLERKERNSILWKLKTP 15898
QY 6 ----- 5
Db 15899 PQTFFKTTGLEGEVEYEFRSAENIYIGKPSVSECYVARDODPPGREALIVTRNSV 15958
QY 6 ----- 5
Db 15959 TLAQKKPTLDGSKITGYIVEKKELPEGRMMKASFTNIIDTTFEYVGLVEDHREFRVIA 16018
QY 6 ----- 5
Db 16019 RNAAGVSESESTGAITANDEVDPRISMDPKYKDTIVVHAGESEFVADIDYKPLPTI 16078
QY 6 ----- 5
Db 16079 QWINGDDELSTARLEIKSTDFATSLVKDAVRDSDGNITLAKKNAVAGERSVTVNVKVL 16138
QY 6 ----- 5
Db 16139 RPPGPEGVVISGYTAECTLAMKPLDGGSDIINYIVERRETSRLVWTVVDANQVTL 16198
QY 6 ----- 5

Db 16199 CKVTKLEGNEXYIRMAVANKYGVGEPLSESPVAVAKNPFVVPDAPKAPAEVTITIKDSMITV 16258
QY 6 ----- 5
Db 16259 WVERPASDGGSEILGYVLEKRDKEGIRWTRCHKRLIGELRLRTVGLTENHDEFRVSAEN 16318
QY 6 ----- 5
Db 16319 AAGLSESPSPSAVQKADPIYKGPNNPKVIDITRSSVFLSMSKPIYDGCCEIQGYIVE 16378
QY 6 ----- 5
Db 16379 KCDVSVGEWTCPTPTGINKTNEVEKLEKHEYNFRICAINNAGGEHADVPGLIIVEE 16438
QY 6 ----- 5
Db 16439 KLEAPDIDLDELKRIINRAGSLRLFPPIKGRPTPEVKMGVDSGEIRDAIITDVTSSF 16498
QY 6 ----- 5
Db 16499 TSLVLDVNRDSCXYITLLENSGTSKSAFYTVAVLDTPSHPVNLKYTEITKDSVSITWE 16558
QY 6 ----- 5
Db 16559 PPLDGGSKIKNYIVEKREATRKSIAAVVTNCHKNSMKIDLOEGCSYFRVTAENEGCI 16618
QY 6 ----- 5
Db 16619 GLPQATPIKVAEVPQPPGKITVDVTRNSVLSWTKPEHDGSKIIQYIVEMQAKHSE 16678
QY 6 ----- 5
Db 16679 KMSKARVKSLOAVITNLJGEEYLFVVAANERGRSDPSLAVPIAKDLVTEPDVKA 16738
QY 6 ----- 5
Db 16739 FSSISVQGDILKIEVPLISGRPKPIITWTKDGLPLKQTRINVTSDSLDTLTSIKETHKD 16798
QY 6 ----- 5
Db 16799 DGGYGITVAVNVGQKTASIEIIVTLKDPDPKPVKFDVSAESITLSMNPPLYTGCOI 16858
QY 6 ----- 5
Db 16859 TNYIVQKBDTITVWDVVSATVARTTIKTKTCTEYQFRIAENRYGOSFALESPIV 16918
QY 6 ----- 5
Db 16919 AQPYKPGPPTPATAISKDSMVIQWHEPVNNGSPVIGYHLERKERNSILWTKVNT 16978
QY 6 ----- 5
Db 16979 ITHDTQKAQNLGIEIEYFRVIAENIVGKASKNSCYVARDODPPGPEIMVKNRN 17038
QY 6 ----- 5
Db 17039 EITLQTKPEYDGSNITGYIVEKRDLPDGRMMKASFTNVIETQTVSGLTEDQRYEFRV 17098
QY 6 ----- 5
Db 17099 IAKNAAGISKSPDSGTPTAKDEVELPRISMDPKRDTIYVNAGETELBADYHGKPLP 17158
QY 6 ----- 5
Db 17159 TTEWLNGKEIESARCEIKNTDFKALLVKDAIRIDGGQVILRASNAVAGSKSPFVNVK 17218
QY 6 ----- 5
Db 17219 LDRPPEGPVOYTGVTSEKSLTWSPLDGGSDISHVVEKRETSRLAMTVVAASEVYT 17278
QY 6 ----- 5
Db 17279 NSLKVTKLEGNEXYIRMAVANKYGVGEPLSESPVLAAMKNPVLPKPKSLEVTINIAKDSM 17338

QY 6 ----- 5
Db 17339 TVCMNRDSDGSETIGYIYERKDRSGIRMIKCNKRITDRLRLVGTGLTEDEHEFEFRSA 17398
QY 6 ----- 5
Db 17399 ENAGVGEPSPATYYKACDVEFRPGPTNAHYDITTKNSTITLAMKPIYDGGSEILIGYV 17458
QY 6 ----- 5
Db 17459 VEICKADEEWOIYTPOTGLFVTRFEISKLTEHOEYKIRVCAINKVGLGATSVPQVTKP 17518
QY 6 ----- 5
Db 17519 EDKLEAPELDIDSELKRGIVVRAGSARHILPFKGRPTPEITWSREGEFTDKVQIEKV 17578
QY 6 ----- 5
Db 17579 NYQLSIDNCORNDAGKIYIKLENSGSKAFYTVKVLDPGPPQNLAVKEVRKDSAPLV 17638
QY 6 ----- 5
Db 17639 WEPIIDGAKVKNYVIDKRESTRKAYANVSSKSTSFVENLTEGAIYFRMAENEF 17698
QY 6 ----- 5
Db 17699 GCVGPVETVDAVKAEPSPPGKVTLTDSOTSASIMMERPEHDGSRVLGYVEMQPKG 17758
QY 6 ----- 5
Db 17759 TEKSIIVAESKVCNAVVTGLSSGOEYQFRKAVNEKGRSDPRVLGVPIAKDLTIQPSLK 17818
QY 6 ----- 5
Db 17819 LPEFTYSIQAGEDLKIEPIYIGRERNISWVKDEPLKQTRVNEETATSTVLHIKEGN 17878
QY 6 ----- 5
Db 17879 KDDGKTYTATNSAGTATENLSYILEKPGPVYGRDEVSADVVISMEPPAYTGGC 17938
QY 6 ----- 5
Db 17939 QISNHYEKRDTTWTWHMYSATVARTTITKTKTGTEYQFRIIPANRRKGSAPLDSKA 17998
QY 6 ----- 5
Db 17999 VIVQYPRKEPGPGTPEVTSISKDMLVOWHEPVNDGTRKIIGYHLEQEKENSILMWKLN 18058
QY 6 ----- 5
Db 18059 KTIPIODTKFKTTGLDEGLEIEFKVSAENIVIGKPKSVSECFVARDPCDPGRPEALVIT 18118
QY 6 ----- 5
Db 18119 RNNVTLMKRPAYDGSKITGYIYEKKDLDPGRMMKASFNNVLETFYVSGLEDQRYER 18178
QY 6 ----- 5
Db 18179 RVIARNAAGNFSEPSDSSGAIITARDEIDAFNANSLDKYKDVIVYHAGETVLEADIRGRP 18238
QY 6 ----- 5
Db 18239 IPDVVMSKDGKLEETAAARMEIKSTIQKTTLVYKDCIRDTGGYIILKLSVGTKSIPIT 18298
QY 6 ----- 5
Db 18299 VKVLDRGPPEGLKYTVTAEKCYLAMNPLODGGANISHYIIEKRETSRLSWTQVSTE 18358
QY 6 ----- 5
Db 18359 VQALNTKVTKLLFGNEXIFRVAHVNTYIGIEPLESPVTACNPFKPPGPPSPPEVSATRK 18418

QY 6 ----- -HSMEXR 11
Db 18419 DSMVVTWARPVDGOTEIEGYLEEKRRDKSGVRRTKONKKTITDRLRLVGTGLTEGHSYER 18478
QY 12 ----- 11
Db 18479 VAAENAGVGEPSVFRACDALYPGPPSNPKVTDTSRSSSVLSAMSKPIYDGAAPYK 18538
QY 12 ----- 11
Db 18539 GYVVEVKEAADDEWTTCTPPTGLQKQFTVTKLENTEYFRICAINSEGVGEPAITLPGS 18598
QY 12 ----- 11
Db 18599 VVAQERLEPPEIADLDRKVVILBASATLRLFTYIKGRPEPEVKKERAGILTRAOIE 18658
QY 12 ----- 11
Db 18659 VSSFTMLVIDNVRDPSGRYMLLENNSGSKAFVNVRLDPSAPVNLITREYKDSV 18718
QY 12 ----- 11
Db 18719 TLSMEPPLIDGAKITNYIYERKRETRKAYATITNCTKTTRIENLOEGSGYFRVLAS 18778
QY 12 ----- 11
Db 18779 NEXGILPAETTEPVKVSPPPLPGRVTLVDYTRNATIKMEKPPSDGSKITGYVEMQ 18838
QY 12 ----- 11
Db 18839 TKGSEKSTCTOVKTLEATISGLTAGEEYFRVAAVNEKGRSDPROLGVPIARDIEIKP 18898
QY 12 ----- 11
Db 18899 SVELPEHTFNKARQKIDVPFKGRPOATVWRRDQOTLKTETRVNVSSTVTYLSLK 18958
QY 12 ----- 11
Db 18959 EASKEDVGTVELCVSNSAGSTVPITVILDRPGPGPIRIDEVSCDSTISWNPPEYDG 19018
QY 12 ----- 11
Db 19019 GCOISNYIEKETTSTWHIVSOAVARTSIKIVBLTGSEYQFRVCAENRGRKSSYES 19078
QY 12 ----- 11
Db 19079 SAVVAEYPPSPGPGTPKVVAHATKSTMLVTWQVYVNDGSRVIGYHLEYKRSILMSK 19138
QY 12 ----- 11
Db 19139 ANKILIDTQMKVSGLDEGLMEYRVAENIAGICKSKCEPVAPARDPCDPGQPEVTN 19198
QY 12 ----- 11
Db 19199 ITRKSVLSKMSRPHYDGAKITGYIVERRELPDGRWLKCNNTNIOETFEVTELTEDORY 19258
QY 12 ----- 11
Db 19259 EFRVFAARNAADSVSEPSSTGPIIVKDDVEPPRVMMQVYKFDVLYVKADEVIAKINADIAG 19318
QY 12 ----- 11
Db 19319 RPLPYISMAKDGIEIEBARREIISTDNHTLLVYKDCIRDTGGYVILTKNVAAGTRSAV 19378
QY 12 ----- 11
Db 19379 NCKVLDRKPPAGPLEINGLTAEKCSLSWGRPOEDGADIDYIYERKREISHLANTICEG 19438
QY 12 ----- 11
Db 19439 ELQMTSCVTKLLKGNEXIFRVTGVNKYVGEPLSAIKALDFYVSPPTSLEITSVT 19498
QY 12 ----- 11

```

Db 19499 KESMTLCWSPESDGSSEISGIIIEERKNSLAWVAVNKKPVYDLAVKSTGLREGCEYEX 19558
QY 12 ----- 11
Db 19559 RYVAENNAAGLSPSESPILRIADPVLPSPPSKPKIVDSKTTITIAWVKPLFDGAPI 19618
QY 12 ----- 11
Db 19619 TGYTVEYKKSDDPTDMKTSIQSLRGTEYTIISGLTGAEYVFRVSKVNGASDPSDDQ 19678
QY 12 ----- 11
Db 19679 IAKEREPEPLFDIDSEMRKTLIVKAGASFTMTVPFRGRPVNVLMSKPPDLRTAYVDI 19738
QY 12 ----- 11
Db 19739 TDSRTSLTIENANRNSGKYTLTIQVLSAASLTLVVKLDPPGPTNTIVDVTESAV 19798
QY 12 ----- 11
Db 19799 LSMVDPENDGAPVKYHIEKREASKAWSVTNNCNRLSYKVTNLQEGAIYIFRVSGEN 19858
QY 12 ----- 11
Db 19859 EFGYGIAPETKEGKITERKSPPEKLGVTISKDSVSLTLKPEHOGSRIVHYVEALE 19918
QY 12 ----- 11
Db 19919 KGOKNWKCAVANKSTHHVSGLENSEYFPRVAENOGSLDRELLPVLKEQLEPPE 19978
QY 12 ----- 11
Db 19979 IDMKNFPSHYVYVAGSNLKVDPISGKPLPKYTLSDGVPLKATMNFENETIAENLTIN 20038
QY 12 ----- 11
Db 20039 LKESVTADAGRHEITANSSGTTKAFINIVLDRPGPTGVVIVSDITEESVTLKWEPPK 20098
QY 12 ----- 11
Db 20099 YDGSQVYTLILKRETSIAVTEVSATVARTMKVAKKLITGEHYQRIAEHRFGISDI 20158
QY 12 ----- 11
Db 20159 IDSACVYKLPYTPPGPSPFWTNTRESITVGMHEPVNSNGSAYVGYHLEMKDRNSIL 20218
QY 12 ----- 11
Db 20219 WOKANKLIVIRTHKVTITISAGLIYEFERYAENNAAGVKSHPSEPVLAIDACEPPRNR 20278
QY 12 ----- 11
Db 20279 ITDISKNSVLSMOOPAFDGSKITGIIVERDLDPDGRWTKASTNTETOFIISGLTON 20338
QY 12 ----- 11
Db 20339 SQYERFVAFARNAVGSISNPSEVGPITCIDSYGVPVLDPLEYEVVKYRAGTSVKLRAG 20398
QY 12 ----- 11
Db 20399 ISGRPAPIEWYKDDKELQTNALVCVENTDILASILIKADRLNSGCEYELKLRNMGASAS 20458
QY 12 ----- 11
Db 20459 ATRVQIILDKPPGPGPIEFETVTAEKITLLMRPADDGAKITHIYERKRETSRVVWSM 20518
QY 12 ----- 11
Db 20519 VSEHLEECIITTTKIKGNEVIFRVAVANKYGIGPLESDSVAKNAFVPGPGIPEVT 20578
QY 12 ----- 11

```

```

Db 20579 KITKNSMTVWMSRPIADGSDISGTFLEKRDKSLGWFKYLKETIRDTROKVTLGTENS 20638
QY 12 ----- 11
Db 20639 YQYRVCAVNAAGQGFSEPSSEFYKAADPIDPPGPAKIRIADSTKSSITLGWSKPYVDG 20698
QY 12 ----- 11
Db 20699 SAVTGYVEIRQGEHEEMTIVSTKGEVARTTEYVSNLKPQVNYFRVSAVNAAGQGEPIE 20758
QY 12 ----- 11
Db 20759 MNEPVQANDILEAPEIDDLARTSVIAKAGEDVQVILPFKGRBPPTVYWRKDEKNLSD 20818
QY 12 ----- 11
Db 20819 ARTSTENTDSSLLTIPOYTRNDQKTYLITENGVEGPKSSTVSXVLYDTPAACQKQVK 20878
QY 12 ----- 11
Db 20879 HYSRGVTLLMDPPLIDGSPILINVIKRDATKRTMSVSHKCSNIFKLIDSEKTPF 20938
QY 12 ----- 11
Db 20939 FFRVLAENEIGIGEDCETTEPYKALEVAPARIDLSMKDSTSVILSWTKPFDGGSYIT 20998
QY 12 ----- 11
Db 20999 EYVERKKGGEOTWSHAGISKTCEIEVSOLKEQSVLEFRVAKNEKGLSDPTVITGPIYK 21058
QY 12 ----- 11
Db 21059 ELITPEVLDSDIPGAQVTVRIGHNVHLELPYKGRKPSISMLKDGLPLKESEFVRSEKT 21118
QY 12 ----- 11
Db 21119 ENKITLSIKNAKHEGKTYVILDNAVCRIVAPITVITLGPSPKPKGPIRDEIKADSVI 21178
QY 12 ----- 11
Db 21179 LSMVDPENDGGEITCYISIEKRETSQTNMKVNCSSVARTTFKVPNLVDAEYQFVRAEN 21238
QY 12 ----- 11
Db 21239 RVGVSQPLVSIIVAKHOFIRIPGPKPVYINWTSDGSLAMDAPVYDGSSEVTGFHYEK 21298
QY 12 ----- 11
Db 21299 KERNSILMOKVNTSPISGREYRATGLVEGLDYQFRVYAENSAAGLSSPDPKFTIAVSPV 21358
QY 12 ----- 11
Db 21359 DPGGTPDYIDVYRETITLKNMPLRDGSKLVGYSIEKOGNERWVRCNFTDVSECQYTV 21418
QY 15 ----- 14
Db 21419 TGLSPGDREYFRILANNAVGTISPPSQSSGIIIMTRDENVPPIVEGPEYFDGLIKSGES 21478
QY 15 ----- 14
Db 21479 LRITKALVOGRPVRYTWFDQVIEIKRMMEITDVLGTSLEFVRDATRDRGVYVEAKN 21538
QY 15 ----- 14
Db 21539 ASGSAKEIKVKVQDPGKVVGPPIRPTNTITGEKMTLLMDAPLNDGCAPIYHIEKRETS 21598
QY 15 ----- 14
Db 21599 RLAVALIEDKCEAOSTAIAKLINGNEYOFVSAVNAKFGVGRPLSDPVVAQIQYTVDPAP 21658
QY 15 ----- 14
Db 21659 GIPEPSNITGNSITTLTWARPESDGSSEIQYILIERREKSTRWVAVISKRPISERTFRVY 21718

```

```

QY 15 -----INP----- 17
Db 21719 GLEGENEFHWAENAGVGPASISRLIKCREPVNPPGPTVVKVTDTSKTVSLEMS 21778
QY 18 ----- 17
Db 21779 KPVFDGMEIIGYIEMCKADIGDHKVAACVKTRTYTDLDQGEYKFRVSAINGAG 21838
QY 18 ----- 17
Db 21839 KGDSCVETGTIKAVRDLTAPBLDIDANFKQTHVRAAGASIRLFAYQGRPTTAVWSKPD 21898
QY 18 ----- 17
Db 21899 SNLSRADIHITDTSFTLJVENCNNDACKYTLJVENNGSKSITFTYKVLDPGPPGP 21958
QY 18 ----- 17
Db 21959 TFKDVTGSAITLMDAPLLDGGARLHHVYVERREASRSMOVISEKTRQIFKVNDLAE 22018
QY 18 ----- 17
Db 22019 VPFYFVSAVNEKVGEPYEMPEPIVATQBPAPRRLDVVDTSKSAVLAMLKPDHDGGS 22078
QY 18 ----- 17
Db 22079 RITGYLLEMRKOGSDPWEAGHTKOLFTYVERLVEKTEYEFVKAKNDAGYSEPREAFSS 22138
QY 18 ----- 17
Db 22139 VLIKPOIEPTADLTGNTQNLITCKAGSPTIDVPISGRPAKVTWKLMEMLKETDRVS 22198
QY 18 ----- 17
Db 22199 ITTKRRTILTYKDSMRGDSGRFELTLENTAGVKTSTYVVIIGRGPYTGPLEVSSVSA 22258
QY 18 ----- 17
Db 22259 ESCVLWMGBPKDGGTEITNYIYERKESGTAMQLVNSVKRTQIKVTHLTKMEYSFRV 22318
QY 18 ----- 17
Db 22319 SSENREGVSKPLESAPLIAEHPEVPSPAPTRPEVHVSAANAMSIRWEEPYHDGSKIIIGY 22378
QY 18 ----- 17
Db 22379 WVEKKEKRTILMYKKNKVCLECNKYVTLGVEGLEIQTALNAGVSKASASRPIMA 22438
QY 18 ----- 17
Db 22439 QNFVDAKPREVTDVTRSTVSLIMSAPAYDGSKVYGIIERKPYSEVDGRMLKCNVTI 22498
QY 18 ----- 17
Db 22499 VSDNFTVTALSEGDTYEFERVLAKNAAGVYSKSESTGPVTCDEXAPPAELDLARLHGD 22558
QY 18 ----- 17
Db 22559 LVYIRAGSDLYLDAVAGKPEPKIWTKGDKELDCEKYSLOYTGKRAVAKFKCDRSDS 22618
QY 18 ----- 17
Db 22619 GKTYLTVKNAAGTKAVSVMKVLDSPGCKLTVSHVQEKTLAMSLPQEDGAEITHY 22678
QY 18 ----- 17
Db 22679 IVERRETSRLNWTIVGECPTLSYVYTRLIKNNETIFRVAANKYGPVVESEPIVARN 22738
QY 18 ----- 17
Db 22739 SFTIIPSPGIPEVYGCKEHIITQWTKPESDGNESINLIVDKREKSLMTVANKDYV 22798

```

```

QY 18 ----- 17
Db 22799 YDTRLKTVLSMEGDDYQFRVAVANAAGNSBPSEASNEISCREBSYTPGPSADRVVDFTK 22858
QY 18 ----- 17
Db 22859 HSISLAWTKPMYDGGTDIVGVLEMOEKDDQWRYHTNATIRNTFTVYDLKMGQKYSF 22918
QY 18 ----- 17
Db 22919 RVAANVWKGMSSEBSIAEIPVERIEIPDELADLKTIVYIRAGASLRMLVSVSGRP 22978
QY 18 ----- 17
Db 22979 PVITWSKOGIDLASRAIIDTESYSLILVDKNRYDAGKYTIEAENSGRKSATVLVKY 23038
QY 18 ----- 17
Db 23039 DTPGPCSVKKEVSRDSVTITWEIPITIDGAPVNNYIVEKREAAHRAFTVYTKSKTL 23098
QY 18 ----- 17
Db 23099 YRISGLVEGTMYFRVLBNENYIGEPCEISDAVLVSVPVPAKLEVDYTKSTVTLAW 23158
QY 18 ----- 17
Db 23159 EKPLYDGGSRITGYVLEACKAGTERMMKVTLKPTVLEHTVLSMEGOYLLFRIRAOEK 23218
QY 18 ----- 17
Db 23219 GVSEPRETVAVTVODLRLFTIDLSMPQKTHVPAGRVELYIPIAGRPPAPASWFEA 23278
QY 18 ----- 17
Db 23279 GSKLSESEVTVETHTKAKLITRETTIRDTIGETYLELKNYGTSTETIKVILDKRGP 23338
QY 18 ----- 17
Db 23339 TGPRIKIDEIDATSITISWPEPELDGAPLGSYVVEORDAHRGMLPVSESVTRSTFKFTR 23398
QY 18 ----- 17
Db 23399 LTEGNEYFRVAANIRREGISYLOSEVIECRSSIRIRPEPETLOIFDVSRDGMLTWYPP 23458
QY 18 ----- 17
Db 23459 EDDGGSQYTVYIYERKEVRADRWVRVNVPTMTIRYRSTGLEGLEEHRVTAIINAGSG 23518
QY 18 ----- 17
Db 23519 KPSRPSKPIVAMDPIAPPKQNPRTDTTRTSVLSAMSVPEDEGSKVTGYLIEQKVD 23578
QY 18 ----- 17
Db 23579 QHEWTKCNTTPKIKREYTLTHLPQGAERYFRVLAACNAGGPEPAEVPQYKVTMLEYPD 23638
QY 18 ----- 17
Db 23639 YELDERYQEGIFVRQGVIRLTIPIKGRPPLCKWTKGQDISRAMIATSETHLVIK 23698
QY 18 ----- 17
Db 23699 EADRGDSGYDLVLENNCKGKRAVYIKVAVIGSPNSPESPLEYDIOVSVSVSWRPPADD 23758
QY 18 ----- 22
Db 23759 GGADILGTYLLEHREVPKAAWTTIDSRVNGTSLVYKGLKENYEHFRVSAENQGISKPLK 23818
QY 23 ----- 22
Db 23819 SEEPVTPKTPLNPPPPSPNPEVLVVTSSVLSMSRKPDDGGSVYGYIERKRTSTDK 23878
QY 23 ----- 22

```

Db 23879 WVRHNKQITTTMTVTGLVDAEYOFRIIAQNDVGLSETSPASEPVYCKDPFDPSPQPG 23938
 QY 23 ----- 22
 Db 23939 ELEIISIKSDSVTLQWKEPCECDGKEIIGVWEYRQSGDSAMKSKNERIKDKOFTIGGL 23998
 QY 23 ----- 22
 Db 23999 LEAFEFERFAENETGLSRPRRTAMSIKTKLTSGEAPGIRKEMKDYTKLGEAQLSCQ 24058
 QY 23 ----- 22
 Db 24059 IVGRPLDINKYRFGKELIQSRKYMSSDGHPTHTLTWTEQEDEGYTTCIATNEVEVE 24118
 QY 23 ----- 22
 Db 24119 TSSKLLQANTPOFHGYPLEKEKYGAVGSTLRHMVITGRVPAMTFHGOKLLQNSENI 24178
 QY 23 ----- 22
 Db 24179 TIENTEHYTHLMKNVQRKTHAGKYKQVLSNVFGTVDAIILDVEIQDKPDKPTGPIVEAL 24238
 QY 23 ----- 22
 Db 24239 LKNSAVISMKPPADDGSGWITNTYVERKECAKEGEMQVSSAISVTTCRIVNLTENAGY 24298
 QY 23 ----- 22
 Db 24299 FRVSAQNTFGISDLEVSIVYIISPEKPGAPGPKPTITAVTKDCVYANKPPASDGGAK 24358
 QY 23 ----- 22
 Db 24359 IRNYYLEKREKQNKWISVTEIEIRETVFSYKNLIEGLEEFYKVCENLGESSEMSISE 24418
 QY 23 ----- 22
 Db 24419 PITKSDVPDIOAPHKEBELNMLNRYQSNATLYCKVTGHPKPIVKYRQKEIITADGLKY 24478
 QY 23 ----- 22
 Db 24479 RIOEFKGYHQILIASYTDDATYQYRATNQGSGVSTASLEVEYPAKTIHLPTLEGMG 24538
 QY 23 ----- 22
 Db 24539 AVHALRGEVVSIIKIPFSGKDPVITMOKGODLIDNNGHYQVIYTRSFSLVFPNGVERKD 24598
 QY 23 ----- 25
 Db 24599 AGFYVCAKRNFGIDQTVELDAVDVDDPPRGYKVSVDVSHDSVNLWTEPASDGSKITN 24658
 QY 26 ----- 25
 Db 24659 YIVEKATFAERMLRVGQARETRTYINLFGKTSYQFRVIAENKFGLSKSPSEPTITK 24718
 QY 26 ----- 25
 Db 24719 EDKTRANNYDEVEDETREVSMTKASHSTFELYEKWIAIDLNGEFGIVHRCVETSSK 24778
 QY 26 ----- 25
 Db 24779 TYAKFYKVGQDOVLVKEISIIINARHNIIHLHESFESMEELVWIFEFTISGLDIFER 24838
 QY 26 ----- 25
 Db 24839 INTSAFELNREIYSYHQVCEALQFLHSHNIGHFDIRPENIITYQRRSSTIKIIEFGQ 24898
 QY 26 ----- 25
 Db 24899 ROLKPGDNFRLLFTAPRYAPEVHQHDVSTANDMMSIGTLVYVLLSGINPLAETNOOI 24958
 QY 26 ----- 25

Db 24959 IENINNAEYTFDEAEKFEISIEAMDFVDRLLYKERKSRMTASEALQHPMLKQIERVSTK 25018
 QY 26 ----- 25
 Db 25019 VIRTLKHRRYYHTLLIKDLNMVVSAAIRISGGAIRSQKVSVAKYKVASIEIGVSGOIM 25078
 QY 26 ----- 25
 Db 25079 HAYGEGGHVKKYCKIENYDQSTQVWTFGVROLENSEKYEITTEDGVALLYVKDITKLD 25138
 QY 26 ----- 25
 Db 25139 DGYRCKVYNDYGEDSSVAELVYKGVREYDYCHRTMKIKRRDTMRLLERPEFTLP 25198
 QY 26 ----- 25
 Db 25199 LYNKTAIVGEBNRFVYTTVHPBPVWTKSGQIKPGDNDKRYTFESDKGLYQLTINSV 25258
 QY 26 ----- 25
 Db 25259 TTDDAEYTVVARNKYGEDSCAKLTVTLHPPTDSTLRPMKRLLANAEQEGSGVCFE 25318
 QY 26 ----- 25
 Db 25319 IRVSGIPPTLWEXKDQPLSIGPNIELIHGLDYALHTRDTLPEDYGYRVATNTAG 25378
 QY 26 ----- 25
 Db 25379 STSCAHLQOVERLRKQKQEFKSEHERHVOKIDKTLMAEILSGTESVPLTOYAKBAL 25438
 QY 26 ----- 25
 Db 25439 REAAVLKRPVAGSTKYVGEFLEIEKKERKLRMPDYPEBRKXQTTIEBQRIKQFV 25498
 QY 26 ----- 25
 Db 25499 PMSDMKMYKIRDOYEMPGLDRVYQKRPKRIRLSRWDQFYVMPLRITDQYRPRKIRPK 25558
 QY 26 ----- 25
 Db 25559 LSQDDLEIVRPARRRTPSPDYDFYRPRRRISGLDISDEBLIIPIDDYLAAMKTEBERLRL 25618
 QY 26 ----- 25
 Db 25619 EEBELGFSAPSPSPRPFELSLRYSQAHVAVEETRKDFRSTYHIPKAASNSTY 25678
 QY 26 ----- 25
 Db 25679 AELREHQAAYRQKQORIMAEREDELLRPVTTQHLSEYKSELDPMSEKESRKK 25738
 QY 26 ----- 28
 Db 25739 RRQREVTETIEIEBEYELSKHAQRESSSASRLRRRSLSPTTIELMRPVSELIRSPQ 25798
 QY 29 ----- 28
 Db 25799 PAEEYEDTERSPPTPERTRPSPPVSSERSLSRPFERSARPDIFSRYSKMAKLTQKT 25858
 QY 29 ----- 28
 Db 25859 SERKYEVLISQDPTLDHAPRITLRMRSHRPOGQNTREILNVQSKPTAEVYHNGVELQ 25918
 QY 29 ----- 28
 Db 25919 ESSKIHNTSGVLLLEILDCHTDQSGTYRAVCITYKGEASDYATLDVGTGYTTIASQR 25978
 QY 29 ----- 28
 Db 25979 RDEEYPRSVFPELTRTEAVAVSSFKTSEMEASSSVREVKSQMTETRESLSYEHSAAE 26038
 QY 29 ----- 28
 Db 26039 MSAALEKSLEKSTTRKITTILARILITRPSMTYEGESARPSCDTDEGPVPTVWL 26098


```

Db 978 SRSRHSASSANISTVGVNDIGTWICERKRKAPSGHPVFRWDFGQKEYATHTQYL 1037
QY 8 ----- 7
Db 1038 SKRSLLVLRISDGHKGLAELLQWLNQARAPNSPVIIVGTHFDVAGESISPOKAEQL 1097
QY 8 ----- 13
Db 1098 QQLIREKFIAPDAEKIGLPRVIDSIEISCTRLNHLNLIYDTAMQLRSPGSKRPM 1157
QY 14 ----- 13
Db 1158 LQKIPASTIALEDIVNVIACNLRAAGDPVLDGEQYRLYTEQRLNHYKSPDAEALQ 1217
QY 14 ----- 17
Db 1218 ATWCHENGVLHYDADATLRDYFLDPQWMLDLAHVYVREINFPFPGVMKLDLQML 1277
QY 18 ----- 17
Db 1278 FRVVOGNGNRSGLRRLILMTYFPGFWSRLITRIADQIIEALRGVYMASODYADF 1337
QY 18 ----- 17
Db 1338 DLRTSLDQDQWNLMTGALTYPIIFIKIWEVPOKTERTOPTDGNRFKLDGIM 1397
QY 18 ----- 17
Db 1398 SDVNLSSSILEYFPLEYVNISQEVVDNEQOLLAETIRPHMSQVAKLLATVDHIDLE 1457
QY 18 AWTXXRGIRPV---GRF 31
Db 1458 DWYPSLCTRFVHTSEGRF 1475

RESULT 14
QY 96077 PRELIMINARY; PRT; 1171 AA.
ID 096077
AC 096077;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Conserved hypothetical protein.
GN B18D24.060.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: CONTAINS 3 WD REPEATS (TRP-ASP DOMAINS).
EMBL: AL513466; CAD11414.1;
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR001680; WD40.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 1171 AA; 130209 MW; DQ8FE35B53232D CRC64;

Query Match 41.0%; Score 64; DB 3; Length 1171;
Best Local Similarity 3.0%; Pred. No. 3.4e+02;
Matches 22; Conservative 1; Mismatches 8; Indels 709; Gaps 6;
QY 1 SRXHXSM----- 8

```

```

Db 397 SRTVRSMTVDRTSNRNPNEFRAPPTPIAATHRIPERIPPEVNPVTSRTANTGPTTM 456
QY 9 ----- 8
Db 457 AYPDATNTNRSKPYISKVHEIYTKYDTRMLDVCGELVCTTGFTRWMSLDGELVMSLS 516
QY 9 ----- 8
Db 517 MGBQKGTAVMFKGAHVDECTKWIGNSGELLEADIGFQRTDSLPNAHGREILRI 576
QY 9 ----- 13
Db 577 YRHFNWTLDDSGALNWGPDESHPTPLNSGPSQSFVRPGRHTESMAYEDELWYATGKD 636
QY 14 ----- 13
Db 637 IVEFKPTVDGKTQFQVLLRALCOQDAGEVIGAVLASDPDMVYFGHSDGKYSYKRKYT 696
QY 14 ----- 13
Db 697 COCINNVSOFKINSIAAGSYLMAGYNNQKAVYDVDTQPMAYKKEWVAHENPVKLIAD 756
QY 14 ----- 13
Db 757 KSSSTKLDRYQVYSLGADNMVFTDALLQEDMLEGEMRKDVEYCEFGNLKVLINTWAG 816
QY 14 ----- 13
Db 817 ASTPHSLRSDSDANPFRLDLOTADSPDIIIFGPOLVLDLKKATAPFLPKKKGSD 876
QY 14 ----- 13
Db 877 QERMSHQYRDWDFLIRSLDDYMGDLVHLHTSPVLGFTICIEFKADLRGRISMSNE 936
QY 14 ----- 13
Db 937 YRRGLGHLGNKGAIVREFLVDDTSLCFINCHLAAGSGANORHNDVAAILDASAPSER 996
QY 14 ----- 16
Db 997 DPAVRIDSYVGGDGTMLIDHELVCVINDIANRIDTMSRDVYVAVKGNLTKLERDL 1056
QY 17 ----- 24
Db 1057 LVAKRRNPGLRARDDELPTFAPYKYVGTGNYDTSEKKRSPAMCDRLYRGRGRIEQ 1116
QY 25 ----- 11
Db 1117 LDYRREAVRVSDBRPVSGRF 1136

RESULT 15
QY 018977 PRELIMINARY; PRT; 4135 AA.
ID 018977
AC 018977;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE Tenascin-X.
GN TN-X.
OS Bos taurus (bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA TISUE-SKIN;
RA MEDLINE-97426436; PubMed-9278449;
RA Eleftherio F., Exposito J.Y., Garrone R., Lethias C.;
RT "Characterization of flexilin, the bovine tenascin-X.";
RL J. Biol. Chem. 272:22866-22874 (1997).
EMBL: Y11915; CAA72671.1;

```

DR HSSP: P02671; 1FZD.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002181; Fibrinogen_C.
 DR InterPro: IPR003961; FN_III.
 DR Pfam: PF00008; EGF_14.
 DR Pfam: PF00147; fibrinogen_C_1.
 DR Pfam: PF00041; fn3; 30.
 DR SMART: SM00181; EGF; 7.
 DR SMART: SM00001; EGF_1like; 2.
 DR SMART: SM00186; FBG; 1.
 DR SMART: SM00060; FN3; 21.
 DR PROSITE: PS00022; FN3_1; UNKNOWN_19.
 DR PROSITE: PS01186; EGF_2; 19.
 DR PROSITE: PS00514; FIBRIN_AG_C_DOMAIN; 1.
 DR EGF-like domain; Glycoprotein.
 KW EGF-like domain; 447376 MW: D903653C3444FA9 CRC64;
 SQ SEQUENCE 4135 AA; 447376 MW: D903653C3444FA9 CRC64;

Query Match 41.0%; Score 64; DB 6; Length 4135;
 Best Local Similarity 0.48; Pred. No. 2.7e+03;
 Matches 17; Conservative 5; Mismatches 8; Indels 4048; Gaps 4;

QY 1 SRXHX----- 6
 DB 56 SRLXEHVEGGEKQVYFTHRINLPASAGCGCPPTPEPVASVQALRYRLLEELVKG 115
 QY 7 ----- 6
 DB 116 LKQCTGGCCPPAAQAQGTGTDIRSLSLHGVFDLSRCACSEPGWGPTCDPEGAEGP 175
 QY 7 ----- 6
 DB 176 PSSPPSPAGSCPDNDQRCVRCVCPFGYTGPGSCSWPSCPGDCHGRKCVQGVCR 235
 QY 7 ----- 6
 DB 236 SGFSGDDCSVRSCPRGCSQGRCEDEGRKVCNPGYSGEDGVSRCPDSCGRKCEMGRCV 295
 QY 7 ----- 6
 DB 296 CNPFGTGDGCVSRCPGSCSQKGRCEDEGRVCNPGYTGDDCGSRTCPMDCGEGRCVDGR 355
 QY 7 ----- 6
 DB 356 CVCMPGTAGEDCSTRICPRRCRGRCEDEGLICDPETSGDDGVSRCPDSCNQRCEDE 415
 QY 7 ----- 6
 DB 416 GRCVCMFGYSGPDGACAPRDCRGRGRCENGVCVCHAGYSGEDGVSRCPDSCRRGRG 475
 QY 7 ----- 6
 DB 476 ESGRCVCMFGYTGNDGSTRACPGDRCGRGRGVDCVRCVNPFGAEDGSRRCPGDGRG 535
 QY 7 ----- 6
 DB 536 RCGDGVSCVGYEGBEDCGKRCRCPGCGGRGCLBGRHCVCDGTEGBEDGVSRCPDSCN 595
 QY 7 ----- 6
 DB 596 RGVCQDGVCTCMBGFAGEDGRLRVCPNSNHRGRGRCENGRCVCDSGYTGPCATRTCPADG 655
 QY 7 ----- 6
 DB 656 RGRGRQGVGVCHVYSGEDGQEEPPASACPGCGPRELCSAGQCVCEGRRGPDCAI 715
 QY 7 ----- 6
 DB 716 QTCGDCRGAGECRGSCVQDGTAGEDGCEVPAIDGMHMLLEETVTEWTRAPGVN 775
 QY 7 ----- 6
 DB 776 DAEIQLPTTEGASPFETARVPSSASAYDQRLAPQGEYQVTVRALRGTNMGPPASKTI 835

QY 7 ----- 6
 DB 836 TTMIDGQDLRVAVTPPTLELMLRQAEYDREVVSAGNORVLEVPSEADGTLT 895
 QY 7 ----- 6
 DB 896 GLMGEVYVTVTAERGRAVSPASIRANTGSSISGLIGATDEPPSGPSTGCAQAPVL 955
 QY 7 ----- 6
 DB 956 QQRPELAELRVLGKDKTGRVAMTAQPDFTFHQLRVPEBGAHELLPGDVRAAL 1015
 QY 7 ----- 6
 DB 1016 VPSPPSPPELTLNGLIPSGEPASAPLIYQIMDKDEKPKPLAPRLGKLTVDVTS 1075
 QY 7 ----- 6
 DB 1076 SILLHWYVEGEFSEFVQYKDRDRPQVVPVEGFORSAISNLVGRKRYKVLGLYK 1135
 QY 7 ----- 6
 DB 1136 RHGPLVAEAKILSQTDPSPVTPPRLGNLWTDPTPDSLHLSMTVPEGQFDSFMYDR 1195
 QY 7 ----- 6
 DB 1196 GRPQVVPVEGPDYSYLIPLDPDHKRYFTLEIANKNRHGLTADGTTAPERKEEPRH 1255
 QY 7 ----- 6
 DB 1256 PPERPLGELIYAGATADSLRLSWTVAQSFDSFVQYKDAQGRQAVPVTDENEVAIP 1315
 QY 7 ----- 6
 DB 1316 SLEPDKYKMLYGLHGRQVGVSVATTAPOEVLDETPATEMEETPSPTSTAPE 1375
 QY 7 ----- 6
 DB 1376 SPEKPLGELMWTGSSPDSLSTWTLPOGHFDSFTVQYRDGQPRVNRVGEDGVTIS 1435
 QY 7 ----- 6
 DB 1436 GLEPDHKYKMLYGRHGRQVGVSVATTAPOEVLDETPATEMEETPSPTSTAPE 1495
 QY 7 ----- 6
 DB 1496 STEADPAEPLGELIYAGATADSLRLSWTVAQSFDSFVQYKDAQGRQAVPVTD 1555
 QY 7 ----- 6
 DB 1556 EVTIGLEPGRKRYKMLYGLHGRQVGVSVATTAPOEVLDETPATEMEETPSPTSTAPE 1615
 QY 7 ----- 6
 DB 1616 DYTNPVGLMTVSEGQDPSFVQYKDRDQSHVVPVAAQDREALVSLGEBRKRYMNV 1675
 QY 7 ----- 6
 DB 1676 GLHGGQVGLSVALTAPVPPDPTPEPVEPRLGELTVTDVTPDSVGLSWTVAEGER 1735
 QY 7 ----- 12
 DB 1736 FLVQYKDRDQSHVVPVATDQREVTPGLEPSRKYKFLFGIQDGKRRSTVSEAKTVGR 1795
 QY 13 ----- 12
 DB 1796 GDASGAPRRLGELMWTDPDPSLRLSWTVPRGHFDSFVQYKDRDQSHVVPV 1855
 QY 13 ----- 12
 DB 1856 TISPLDSGRKRYFLVYGLGKRRHGLTTEGTTETRAVDAGTKRPSKPRLGELIY 1915
 QY 13 ----- 12

Db 1916 VTSNSVGLSWTVBEGHFDSPVLOYRDRDQOPQVVPVEGSRREVSVCIDPARRYKLLLYG 1975
QY 13 ----- 12
Db 1976 LSRDRKRGVISAIAVTERAPREBEIKAEPAESPASEPLGEVTLGEAAPHSLRLSWTAT 2035
QY 13 ----- 12
Db 2036 EGEFDSFEVQYTDENGQOQEVNVGSDQHDITISDLESBHRVLSLYGFHDGQRYGPAHIE 2095
QY 13 ----- 12
Db 2096 AMTAREDEDEPSESLSTTQTPSTAVPERHIKPRIGELAVDTTPDLSLSWTVPEGOFD 2155
QY 13 ----- 12
Db 2156 HFLIOYKNGDQOPKVVVRVPGDEVTYISGLEPDHKYKMNLYGFHNRORMGFVSVIGVTTA 2215
QY 13 ----- 12
Db 2216 EEPSPTEMEETSPTEMEETSPTEBNTENVPERPEPFLBELTVTGSSPDSLSWTV 2275
QY 13 ----- 12
Db 2276 PQGHFDSFTVOYKNGDQOPKVVVRVPGHDKGVYISGLEPDHKYKMNLYGFHDQRVGPVS 2335
QY 13 ----- 12
Db 2336 IGVTTAEETPSPTEVEETPSPTEBNTAVPERPEPVIGELMTVGSSPDSLSWTVPOG 2395
QY 13 ----- 12
Db 2396 HEDFTVOYKGRDQPVVVGGEETEYVGELEPGHKYKMNLYGLHGRVGPVSTVAMT 2455
QY 13 ----- 12
Db 2456 AREEPASPPKPOLGELVYTDATPDSLSWTVBEGQDFHLYQKNGDQOPKAVRVP 2515
QY 13 ----- 12
Db 2516 GDEGVTISGLEPDHKYKMNLYGFHDQRVGPVSTVGLTVSEKQEMTPAPDTLPAAPE 2575
QY 13 ----- 12
Db 2576 PPIKPRIGELVYTDATPDSLSWTVBEGQDFHLYQKNGDQOPKAVRVPDEDEVTIS 2635
QY 13 ----- 12
Db 2636 GLEPDHKYKMNLYGFHDQRVGPVSVIGVTTAEETPSPTEMEETPSPTEVEETPGFTEV 2695
QY 13 ----- 12
Db 2696 EETSPTEPSTEAPEREPEPLGELVYTGSSPDSLSWTVPOGHFDSFTTIQHKGRDGPQ 2755
QY 13 ----- 12
Db 2756 VVRVGEETEYVIGLEPGRKRYKMNLYGLHSGORVGPVSTVDPQEVVEETPSPTEPS 2815
QY 13 ----- 12
Db 2816 TEAPEREPEPLLDLYTVTGSSPDSLSWTVPOGHFDSFTTIQYKGRDQOPVVRVGEETE 2875
QY 13 ----- 12
Db 2876 VTIGLEPGRKRYKMNLYGLHGRVGPVSTVGTAPDYDAMTQTPSTSVPERPTKPRLG 2935
QY 13 ----- 12
Db 2936 EYTVTDATPDSLSWTVBEGQDFHLYQKNGDQOPKAVRVPDEDEVTYISGLEPDHKY 2995
QY 13 ----- 12

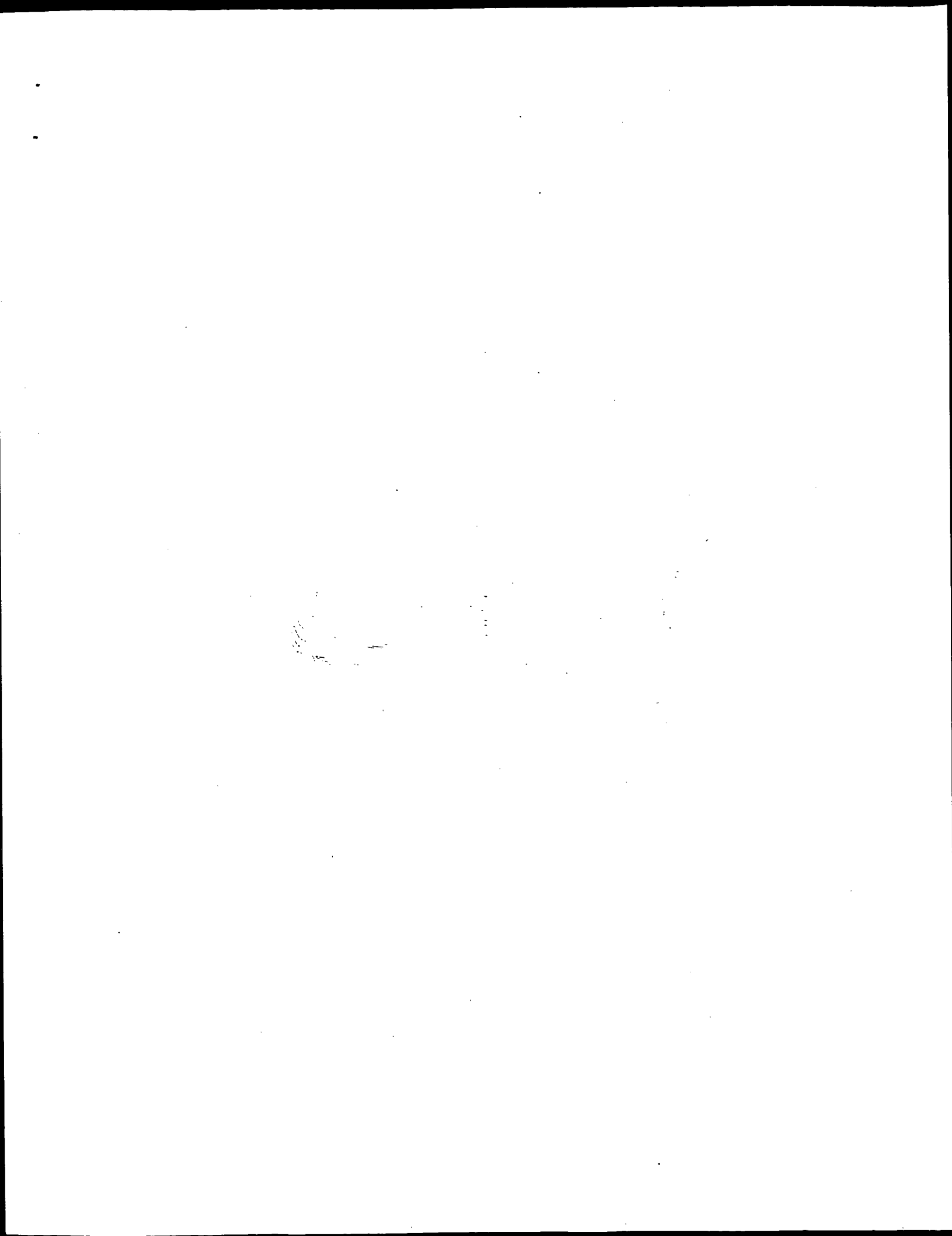
Db 2996 KMNLYGFHDQRVGPVSVIGVTTAEETPSPTEMEETPSPTEMEETPSPTEBNTAVPERP 3055
QY 13 ----- 12
Db 3056 EEPFLGELTVTGSSPDSLSWTVPOGHFDSFTVOYKGRDQOPVVRVGEETEYVIGLE 3115
QY 13 ----- 12
Db 3116 PGRKRYKMNLYGLHGRVGPVSTVGTASLTTERPLAPRLGELAVAVVTSPTARLSWTV 3175
QY 13 ----- 12
Db 3176 QGPFDSFLYQKDVQOPQAVPAVDLREVTVSLAPGRKYKFLFGLRDEKRGHPVSAD 3235
QY 13 -----PDINPA----- 18
Db 3236 AKTLPTKPRAPRLGELVYTDVTPGVSGLSWTVBEGFDSFMQYKDRDQOPRVPAADQ 3295
QY 19 ----- 18
Db 3296 REYTVGLEPNRKYKFLYGLVGRKRLGPIISAGSTAPLEKEROPPRRLGELVYDDETPN 3355
QY 19 ----- 18
Db 3356 SLRLSWTVAGRPDSFVQYRGTDGQPRMVPAADQREFTEVGLBPGRKYKFLYGLLGG 3415
QY 19 ----- 18
Db 3416 QRLGPASVIGMTAPEEDTAPPMHATAPKPEPGPRGLVLAIRDVSPPDSLRLSWVQOP 3475
QY 19 ----- 18
Db 3476 FDSFVQYQDIDQPOALLVGDQNKVLVSGLEPSTSYEFFLYGLHEGRKLGVSABGTT 3535
QY 19 ----- 18
Db 3536 GPVPAGOTPEGPRLSHLSVTDVTTSSLRLNWEAPPEAFDSFLRGVSPSTLEPOLR 3595
QY 19 ----- 18
Db 3596 PILQRELIVPGRRSANVLADLHPGILYITLYGLGPKHADSIOGTARTLSVLESPDL 3655
QY 19 ----- 18
Db 3656 QFSEIETSARVSWPPTSRVDGFKYSQOLADGGEPOSQVVDGRTQKLEGLIPGAQYEV 3715
QY 19 ----- 18
Db 3716 VVSVRGFESEPLTGFLLTVPDGPHLRALNLDESALLHMKPQTPVDYDVKTABGA 3775
QY 19 ----- 18
Db 3776 PSLQASAPGSADVPIQGLVYHTNTATLRLGRLGPNTPSPASITTTGLEARODLEAEV 3835
QY 19 ----- 18
Db 3836 TPRTALLTWTABEVSEPTGYLSFNTPGQTOEILLPGVTSHQRLGLPSPYSTWLRAM 3895
QY 19 ----- 18
Db 3896 WGDSTPPVSTSTFTTGLRIPEPPROCGEEMONGVSTRTTIFLNGNRRLNFCDMET 3955
QY 19 ----- 18
Db 3956 DGGWLVFQRRMDKTDIFRWOMEDYAHGFNISGEFWLGNELHSLTKAGDYSLRVDLRA 4015
QY 19 ----- 18
Db 4016 GDEAVFAQYDSFQVDSADERYRLHLEGYHGTAGDSMSYHSGVSFSGARDRDPNNLLISCAV 4075
QY 19 -----WXXHGI-----RPVGR 30
Db 4076 STRGAMWYRNCYANLNLGLYSTVDHOGVSWYWKGFDFPSVFTTEMKLPRSPYRPLGR 4133
QY 19 ----- 12

Thu May 1 15:54:12 2003

seq74-plus-73.ispt

Page 57

Search completed: May 1, 2003, 14:49:24
Job time : 131 secs



GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:45:32 ; Search time 43 Seconds
(without alignments)
73.778 Million cell updates/sec

Title: SE074-PLUS-73

Perfect score: 156
Sequence: 1 SRXHXSMEXRPPDINPAMYXXRGIRPVGRFX 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	94.2	83	2 JC7607	prolactin-releasin
2	69	44.2	15281	2 S41309	cyclosporin synth
3	67	42.9	7463	2 T36248	CDA peptide synthe
4	66	42.3	26926	1 T38344	titlin, cardiac mus
5	64	41.0	4135	2 T42629	tenascin-X - bovin
6	63	40.4	3063	2 S55505	fatty-acid synthas
7	62	39.7	1762	2 T03222	probable polyketid
8	62	39.7	2476	2 T34022	zonadhesin - pig
9	62	39.7	2611	2 T14591	actinomycin synthe
10	62	39.7	2870	2 A35548	319k protein ndvb
11	62	39.7	5825	2 T12117	polyprotein - fava
12	62	39.7	8363	2 T30226	polyketide synthas
13	61	39.1	1059	2 T20802	hypothetical prote
14	61	39.1	5138	2 B96695	hypothetical prote
15	61	39.1	6831	2 A88852	protein unc-22 [lm
16	61	39.1	6839	2 S57242	twitichin [simlari
17	61	39.1	7160	2 T27935	hypothetical prote
18	60	38.5	276	2 D70817	hypothetical prote
19	60	38.5	873	2 B41054	fasciclin II pi-11
20	60	38.5	962	2 H69157	exonuclease ABC c
21	60	38.5	161776	2 T52022	Munc13-2 - rat
22	60	38.5	2028	2 T52022	1-phosphatidylinos
23	60	38.5	2051	2 P96529	probable phosphati
24	60	38.5	3869	2 A48205	All-1 protein +GTE
25	60	38.5	3968	2 A44265	trithorax homolog
26	60	38.5	4391	2 A38096	perlecan precursor
27	60	38.5	4568	2 T08030	dynamin beta heavy
28	60	38.5	6420	2 T30283	polyketide synthas
29	60	38.5	8243	2 T31307	type I fatty acid

30	59	37.8	664	2 F83376	conserved hypothet
31	59	37.8	866	2 A95272	hypothetical prote
32	59	37.8	940	1 BVCEUA	exonuclease ABC c
33	59	37.8	940	2 H91258	excision nuclease
34	59	37.8	940	2 D86099	excision nuclease
35	59	37.8	941	2 AEL017	excision nuclease
36	59	37.8	947	2 AF0040	excision nuclease
37	59	37.8	1138	2 A48944	excision nuclease
38	59	37.8	1555	2 S38758	paraportel crystal
39	59	37.8	1779	2 T31085	amylo-alpha-1,6-gl
40	59	37.8	1807	2 T30940	xylosylase - Caldice
41	59	37.8	2273	2 T09083	vitellogenin - Pim
42	59	37.8	3341	1 A42996	hemagglutinin/hemo
43	59	37.8	3623	2 T08618	genome polyprotein
44	59	37.8	4466	1 S17653	intrinsic factor-B
45	59	37.8	5107	2 T29144	dynein beta heavy
					partial CDS - Caen

ALIGNMENTS

RESULT 1
JC7607
prolactin-releasing peptide - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: JC7607
R.Yamada, M.; Ozawa, A.; Ishii, S.; Shibusawa, N.; Hashida, T.; Ishizuka, T.; Hosoya, Biochem. Biophys. Res. Commun. 281, 53-56, 2001
A:Title: Isolation and characterization of the rat prolactin-releasing peptide gene:
A:Reference number: JC7607; MUID:21092785; PMID:11178959
A:Contents: Spleen
A:Accession: JC7607
A:Molecule type: DNA
A:Residues: 1-83 <YAM>
A:Cross-references: DDBJ:AB040612; DDBJ:AB040613
C:Comment: This peptide induces archidonic acid metabolite release from rat anterior release, and stimulation of ACTH secretion from the pituitary.
C:Genetics:
A:Gene: Prpr
A:Introns: 33/1
Query Match 94.2%; Score 147; DB 2; Length 83;
Best Local Similarity 83.9%; Pred. No. 1.4e-09;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 SRXHXSMEXRPPDINPAMYXXRGIRPVGRF 31
DB 22 SRXHXSMEXRPPDINPAMYXXRGIRPVGRF 52
RESULT 2
S41309
cyclosporin synthetase - cyclosporin fungus
C:Species: Tolypocladium inflatum (cyclosporin fungus)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 03-Nov-2000
C:Accession: S45487; S41309
J.Weber, G.; Schoergerdorfer, K.; Schneider-Schaefer, E.; Lettner, E. Curr. Genet. 26, 120-125, 1994
A:Title: The peptide synthetase catalyzing cyclosporine production in Tolypocladium n
A:Reference number: S45487; MUID:95094306; PMID:8001164
A:Accession: S45487
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-15281 <WEB>
A:Cross-references: EMBL:Z28383; NID:9440168; PIDN:CA82227.1; PID:9440169
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1993
A:Note: Only a part of the translation is shown
A:Note: The source is designated as Tolypocladium inflatum
C:Genetics:
A:Genetic code: SGC3
A:Superfamily: cyclosporin fungus cyclosporin synthetase; acetate-CoA ligase homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:532-1008/Domain: acetate-CoA ligase homology <ACLI>
 F:1029-1096/Domain: acyl carrier protein homology <ACP1>
 F:1618-2069/Domain: acetate-CoA ligase homology <ACLI2>
 F:2527-2594/Domain: acyl carrier protein homology <ACP2>
 F:3115-3553/Domain: acetate-CoA ligase homology <ACLI3>
 F:4014-4081/Domain: acyl carrier protein homology <ACP3>
 F:4601-5057/Domain: acetate-CoA ligase homology <ACLI4>
 F:5506-5573/Domain: acyl carrier protein homology <ACP4>
 F:6094-6546/Domain: acetate-CoA ligase homology <ACLI5>
 F:7003-7070/Domain: acyl carrier protein homology <ACP5>
 F:7591-8042/Domain: acetate-CoA ligase homology <ACLI6>
 F:8063-8130/Domain: acyl carrier protein homology <ACP6>
 F:8658-9157/Domain: acetate-CoA ligase homology <ACLI7>
 F:9558-9625/Domain: acyl carrier protein homology <ACP7>
 F:10146-10586/Domain: acetate-CoA ligase homology <ACLI8>
 F:11055-11122/Domain: acyl carrier protein homology <ACP8>
 F:11635-12106/Domain: acetate-CoA ligase homology <ACLI9>
 F:12127-12194/Domain: acyl carrier protein homology <ACP9>
 F:12715-13159/Domain: acetate-CoA ligase homology <ACLI10>
 F:13623-13690/Domain: acyl carrier protein homology <ACP10>
 F:14213-14676/Domain: acetate-CoA ligase homology <ACLI11>
 F:14698-14765/Domain: acyl carrier protein homology <ACP11>
 F:14698-14765/Domain: acyl carrier protein homology <ACP11>
 F:1060,2558,4045,5537,7034,8094,9589,11086,12158,13654,14729/Binding site: phosphopantet

Query Match 44.2%; Score 69; DB 2; Length 15281;

Best Local Similarity 0.2%; Pred. No. 8.2e+03;

Matches 20; Conservative 3; Mismatches 8; Indels 11650; Gaps 5;

QY 1 SRX-----4
 Db 1351 SETHOVSFAVLLAFAFRAHFRLTGSDNATIGVPSANRNPPELVIGFVNTQICIRITI 1410
 QY 5 -----4
 Db 1411 DENDNFESLYRQVSTTTAODNDVPPFEQVYSSILMPSSSRDASRNPVLQMLFALHGGOD 1470
 QY 5 -----4
 Db 1471 LFKIQEGTEEEVPIPEEYTRFDIEFHLYOGASKSLGDIIFAADLFEAETIRGVSVFOE 1530
 QY 5 -----4
 Db 1531 VLRGCIQOPQPIQMPPLTDGIPLEIRMGKLMHVKTDYPRNKSVDVFOQOVRLSAETA 1590
 QY 5 -----4
 Db 1591 VIDSSRMSYAEIDQRSDQVAAWLRQRLPAETFAVLAPRSCAVIALFGILKAGHAYL 1650
 QY 5 -----4
 Db 1651 PLDVNVPAARLAILAEVNGEKVLVLAGESPSPGOSPEVSIVRIADATSPAGHASLRDG 1710
 QY 5 -----4
 Db 1711 KSKPTAGSLAVYITFSGTGKPKGVIIEHRYLRLVKOTNILLSLPPAQTFMAHMSNLA 1770
 QY 5 -----4
 Db 1771 FDSIWEFTALLNGSLVLCIDFTLLDAQALEFLREHINIALFPALLKQCLTDAA 1830
 QY 5 -----4
 Db 1831 TTKSIDLLVVGDRDLDTADALAKALVNSEYVNAVPTENTVSTLYSIADTERFVNGVP 1890
 QY 5 -----4
 Db 1891 IGRVNSNGSVYMDONQOLVPLGAVGELVYTGDLARGITNPALDSDFVDVIANGQLLR 1950
 QY 5 -----4
 Db 1951 AVRTGDRARYRPKDGQVEFFGRMDHQVKGRIELAEVEHALLSAGVHDAAVYVNSQGE 2010
 QY 5 -----4

Db 2011 DNGEVAVFITADNNTLQEAQSSNQVQWESHFETTAVADITAIIDQNTIGRDFTSWTS 2070
 QY 5 -----4
 Db 2071 MYDGLIDKREMOEWMLDTRTFLDQAGAHVLEIGTGVMVLFNLGQGLKSYIGLEPS 2130
 QY 5 -----4
 Db 2131 QSAQVFNKAAQTFPGLEGQAQVHVGTMPTGRSLSPOLIYINSVAQFFPSREYLAEV 2190
 QY 5 -----4
 Db 2191 VEALVRIPIGVRIPIFGDMRTYATHKDELVARAVHTNGSKVTRSKVOEVARLEEBEILL 2250
 QY 5 -----4
 Db 2251 VDPAFSTLSLESISEIEHVEILLPKMKVNNELSSYRGAVLHIRHNOQSHIRINA 2310
 QY 5 -----4
 Db 2311 ESWIDFASQMDROGLARLLEKKNKAESIAVFNIPIYSKTIYERHIAKSLADHDGDTHS 2370
 QY 5 -----4
 Db 2371 SIDGVAMISAAREKASQPSLDVHDVQLAEDAGFRVEVSWARQSGALDVFHHFQP 2430
 QY 5 -----4
 Db 2431 TENESRALVDFPDIYKQQAQARSILNRPLOVESRRIEQAQREQLVLLPAYMIPARIYVL 2490
 QY 5 -----4
 Db 2491 QNMPLNTSGVDRKRELILRAKVTAARTPSESELVAPRDSIEAILICEKFKDVLGVEVGTION 2550
 QY 5 -----4
 Db 2551 FFNVGSHSLATKILARLSQNAQIAVKDIFDRPYADIATIQDTEHNPLPTSYT 2610
 QY 5 -----4
 Db 2611 GPVEQSFAGRLMFLDOLNNGATWYLMFPAVRLRGLVVSALAALLAEERHETLRTEF 2670
 QY 5 -----4
 Db 2671 IEQEGIMQVLIHPAPKELAVIDVSGEESTIQKILEKQTPFNLASERFRLALLKTG 2730
 QY 5 -----4
 Db 2731 EDEHILSTVNHHAISDGSVYDIFQOEIGQFYSAILRGHDPLOAIAPLSIQYRDFATWQRQ 2790
 QY 5 -----4
 Db 2791 IFQVAEHRQLAVYTKQADANKPAELLTDKRRPMLSGRAGELPVVYDGLIYEKLQDFCR 2850
 QY 5 -----4
 Db 2851 IRQVTAFTVLLAARAHYBMTGHEDATIGTPIANRNPBELGLIGFVNTQCMRTIVDV 2910
 QY 5 -----4
 Db 2911 EDSEFTLVHQRRETTIAAHANODVPEQIVSNILPSSSDTSRNPVLQMLFALHSQNLGR 2970
 QY 5 -----4
 Db 2971 VRLEGIEEELISIAETTRDIEFHLYQEAERLNGSIYAADLFVPETIOSVITIFQGIIG 3030
 QY 5 -----4
 Db 3031 KGLCEPDMVPAASMALDGLSLRSTGLHPQOTDYPQDASVQVIFKQYAAVNPDIAYVD 3090
 QY 5 -----4


```

Db 3091 ESTRLSYADLRKSDOVACMLSRGIAPETFFVAILAPRSCETIVAILGVKANLAVLPID 3150
QY 5 ----- 4
Db 3151 VNPASRLAEILSEVSGSMLVVGAEPIPEGMAEAEITRIETILLADAKTDINGLAAG 3210
QY 5 ----- 4
Db 3211 PTAASLAVYIFTSSTGRPKGVNVEHRIIVRLTKQTNTSKLPESPFMAHISMIAFASV 3270
QY 5 ----- 4
Db 3271 WEFTTILNGTLVCIDYFTLESTALEKVFEDQRVNVALLPALLKQCLDNPALVKTLL 3330
QY 5 ----- 4
Db 3331 SVLYIGDRLDASDAKARGLVQTAFNAGPTENTVMSTIYPIADPPLNGVPIGHAYS 3390
QY 5 ----- 4
Db 3391 NSGAFVMDNOQITPPGAMGELIVTGDLARGYTSLSNTGREINVDIDEGQVAVRTGD 3450
QY 5 ----- 9
Db 3451 RVRYRRKDLQIEFFGRIDHGVKIRGRIRPAVEYFALLSHDLVTDAAVYTHSOENODLEM 3510
QY 10 ----- 9
Db 3511 VGFVAARVADVREDESSNQVEMQTHFDSIAYADITTTIDQSLGRDFMSWTSWYDGLK 3570
QY 10 ----- 9
Db 3571 KSGOMELMDTMRSLDSDPGHVLVGTGTGVNLFNLGREGLOSGVLEPSPSATAFV 3630
QY 10 ----- 9
Db 3631 NKAKSPGLEDRIRYEVGTANDIDRLGDDLHAGLVVNSVAGYFSPQDYLAQVADLT 3690
QY 10 ----- 9
Db 3691 VPGVERIFFGDMRSHAINRDLVARAVHALGDKATKAEIQREVMESEDELLVDPAF 3750
QY 10 ----- 9
Db 3751 TSLTTOENIKHVELLPKRRATNELSSRYAAVLHNDLAPRAKVSFGAWYFATKM 3810
QY 10 ----- 9
Db 3811 DRDALIRLKGTSIDHIAIANIPNSKTIVERTICESYVDLGDAKDSNDRYSWLSAANS 3870
QY 10 ----- 9
Db 3871 NAVKVASLSAIDLVDIAOAGFVEISCAROWSONGALDAVFHHLCPSSQSHVLIDFLT 3930
QY 10 ----- 9
Db 3931 DHQGRPEALTNHPLHRAOSRVERQIRERLOTLPLAYMIPAQIMVLDKPLPNANGKVD 3990
QY 10 ----- 9
Db 3991 KULTORAGTVPKAKOVSAPVAPRTETIERVLCQEFSDVLDGIMENFTDLGHSIMATK 4050
QY 10 ----- 9
Db 4051 LAARISRLTETHVSKVEIFDHPRVCDLVIVOGSAPHDPPIVSTKYTGHPQSFAGRLM 4110
QY 10 ----- 9
Db 4111 FLDQINFGATWYMLAVLRGAMNHALTAALLAERHHELRTRTFYQNGVGMOKVNP 4170
QY 10 ----- 9
Db 4171 VVETLRIIDLSNGDDYLPILKKEQTAPFHELETPGWMVALLRLGPGDYILSVVHHII 4230

```

```

QY 10 ----- 9
Db 4231 SDGWSVDVLFQELGQFYSPAVKCHDPLSQOTPLPIHYEDFALMOKKPTQSEHEROLQY 4290
QY 10 ----- 9
Db 4291 VEQLVSAPELLDLPRLSILSGQAGMSVTLEGATKNLEFCRVHRTSFVLLAAL 4350
QY 10 ----- 9
Db 4351 RAAHRLTGSEDAITGTPIANRNPBEQLIGFVNTQIRITVNEDETFESLVQOVRST 4410
QY 10 ----- 9
Db 4411 ATAAFAHODVPRKIVSTLLPGSRDASRNPVLVOLMAFVHSOKNIGELKLENABSEVVPTE 4470
QY 10 ----- 9
Db 4471 ITTRFDEFFHLEQDDKLEGSILYSTDLFEAVSQSLSTVFQELLRGLNGPDPVISTLP 4530
QY 10 ----- 9
Db 4531 LODGIYDLQROGLDVQKTEYPRDSSVVDVHEQVSNPDSIALIHGSEKLSYAQDRES 4590
QY 10 ----- 9
Db 4591 DVAARMLHRSSDPTLAVLAPRSCETIIFLGILKANLAVLPDYKAPARIDATYSS 4650
QY 10 ----- 9
Db 4651 LPGNKLLIGANVTPPKLOEAIIDFVPIRDTFTLLTDGDTIERPSAQSLAYMFT 4710
QY 10 ----- 9
Db 4711 SSGTRPKGVNQHNYIYRLVKNSNVAKOPAAARIAHISMIAFADASSWEIYADLNGA 4770
QY 10 ----- 9
Db 4771 IVCADYFTTIDPQALQETFOHEIRGAMLPPSLKQCLVQAPDMISRDLIFAAQDRFSS 4830
QY 10 ----- 9
Db 4831 VDALQARLVSGVFNAGPTENTILSTIYVNAENDESVNGPIGSAVNSGAYIMDKQ 4890
QY 10 ----- 9
Db 4891 QLVPAVGELVYTGDLARGYMDPKLDADRFIQITVNGSEQVRAVYRTGDRVYRPRDFQ 4950
QY 10 ----- 13
Db 4951 IEEGRMDQOKIRINGHRIEPAVEQAFNDGFVEDVAIVIRTPENQREMAVATLAKDN 5010
QY 14 ----- 13
Db 5011 SAREEATTOIEGWEAHEGAGYANIEIESEALGYDFMGTSMYDTEIDKEMREMLN 5070
QY 14 ----- 13
Db 5071 DTMRLDGKPRAGVLEVGTTGIMENLGRSGLERTIGLEPAPSAEYVNNAAKSPFG 5130
QY 14 ----- 13
Db 5131 LAGRAEVHGTAAVGTLOGLTSDMAVINSVAQYEPPEYLAEITKSLVQPGMKRIYLG 5190
QY 14 ----- 13
Db 5191 DMRSWANRDPAAAPAYSLADNASKDRVRQKMELEKEBELLYDPAFTALASQLODR 5250
QY 14 ----- 13
Db 5251 IQHVEILPKRKATNELSSRYAAVLHISDEPLTIYKIDPEAMINFGSRLTREALAQVL 5310

```

QY 14 ----- 13
Db 5311 KENENAESVATSNIPYKTYVERHIVRSLOQEDANAPESMDGDMISAVTRAQOCHTL 5370
QY 14 ----- 13
Db 5371 SASDLFDIAEDAGEFVEVSMAROSHOGALDAVPHHLKATEDSRVLKFPPTDHQGRPLK 5430
QY 14 ----- 13
Db 5431 SLTNOPLPAOSRRARELLIREGLQTLFPPYMPISQITLIDMPLNANGKVDRELARRAK 5490
QY 14 ----- 13
Db 5491 ITOKSPVEDIYPPRNSYEATYCKGFTDVLGVEGITDNFNLGHSIMATKLARLGRQ 5550
QY 14 ----- 13
Db 5551 LNTIRISVDVEDQPVVADLAIVIQNSAPHEPIKPADYTGVPQSFAGRLMFLDQLVNG 5610
QY 14 ----- 13
Db 5611 ATWYLMPLGIRLHSGRLVDALATATASLEQRHEPLRTTFHEDEGVGVQVQDHRPKDLRI 5670
QY 14 ----- 13
Db 5671 IDLSTOPDAYLAVLKEHQTLFLDLATEPGWRVALIRLGEHEHLSIVMHHIISDGMSVE 5730
QY 14 ----- 13
Db 5731 VLEDENHRYPSALNOOPMEOILPLIQYRDPAAOQTEBOVAEHQOLDYWTBHLADS 5790
QY 14 ----- 13
Db 5791 TPARELTDLPRESILSGRANELPLIEGRLDKLAFRCVHQANPFIYLLAALRAHYL 5850
QY 14 ----- 13
Db 5851 TGAEDATIGTPIANRNRPELEMMIGFVNTQCMRAIEENDNFESLYRVRSTATSAFAN 5910
QY 14 ----- 13
Db 5911 QDVPESTIVSSLLPGSRASRNPLVQVLLAVHSOODLCKLLEGLRDAVDAISAISTRDV 5970
QY 14 ----- 13
Db 5971 EFHLFEHADRLSGSVLYAKELFKLTIESVSVFLETLRALDQPLFPLAVIPLTDGVE 6030
QY 14 ----- 13
Db 6031 IASKGLDVPRTDYPRDANIVEFQOHRATPDATAVKDATSILTYAQDOOSDLAIWL 6090
QY 14 ----- 13
Db 6091 SRHNMPELIVGLAPRSCETIAMEGIMKANALVLPIDINSPARLRLISLAVDGNKLV 6150
QY 14 ----- 13
Db 6151 ILGSGVTAPEQENPEVEAVGIOELLAGTGLDKTQGSNARPSATSLAVIYFTSGSGKRG 6210
QY 14 ----- 13
Db 6211 VWEHRSVTRLAKPSNVISKLPOGARVAHLNIAFDASIMEIATTLNGATLVCLDYHTV 6270
QY 14 ----- 13
Db 6271 LDCKRLKEVERESTIVYTLMPALLKQCVAFETIHALDLITYGGDRVSGHDAMRARS 6330
QY 14 ----- 13
Db 6331 VRIKMGSGYPTENTVISTIVEVDADEMFVNGVPIGKTVNSGAVYMDRNOQLVPSGVG 6390
QY 14 ----- 13

Db 6391 ELVYTGDLGARGYTDPSLKNKRPIYITVNGESIRAVRTGDRVRYRPHDLQIEFGMDQO 6450
QY 14 ----- 13
Db 6451 VKIRGRIBEPGEVESALLSHNSVODAAVVICAPADQDSGAEVAVPAAANTEDEDTQEEE 6510
QY 14 ----- 13
Db 6511 ANDVOGWETHFETAAYSEVKIDROSEVGNDFMGWTSMTDSSEIDKTDHMEWLNQTMRI 6570
QY 14 ----- 13
Db 6571 LDAREPGHVEIGTGMWVFNLAQCPGLQGYGFEPSKSAQFVNDAAQSEPALKGRS 6630
QY 14 ----- 13
Db 6631 IVHVGATDINKAGPIQPLVINSVAQYPPPEYLFVVEALVOIPSEVRIVFGDMRTN 6690
QY 14 ----- 13
Db 6691 AINRDFVASRALHTLGEKANKRLVRQMIYELEANEELLTPAFPTSLTRLGEKIKHYE 6750
QY 14 ----- 13
Db 6751 ILPTMKATNELSKYRYAAVLHVRSREOSTIHQVSPNAMIDFADGLDQTLINLKEH 6810
QY 14 ----- 13
Db 6811 KDAGTVAIGNIPYKTIIVERFVKSLSEDDMEGNSLDGSAMVAAPRAAOSCSLDM 6870
QY 14 ----- 13
Db 6871 DVKEIAQAGYQVEVSMAROSQNGALDAIFHHEPPEKGAFTLIEPTDYGRNVTLT 6930
QY 14 ----- 13
Db 6931 NRPLNSIOSRLQTOIREKLQTLPPYMIPSRLVMDMPVNNNGIKDKELVRRATVAP 6990
QY 14 ----- 13
Db 6991 KPRSAATRVAPRNEIBAILRDEFEDVLGTEVSVDNFPLDGHSLMATKLAAVSRRLDA 7050
QY 14 ----- 13
Db 7051 HISIKVFDQPVLAADLAASIORESAPHEPIQRPYTPAQSFAGRLMFLDQNLGATW 7110
QY 14 ----- 13
Db 7111 YLMPALIRIGQLRVAALSAALFALERRRHETLTTFEESDGVGVQIVGEARNSDLRVHDV 7170
QY 14 ----- 13
Db 7171 STGDDGEYLEVLARQOTVPRDLSSEPGWRVCLVKTGEEDHVLISVMMHIIYDGMVSVDILR 7230
QY 14 ----- 13
Db 7231 GELGQFYSALRGQDPLLANPLPIQYRDPAAQORAKOVEEHOROLGYWSQOLVDSFPA 7290
QY 14 ----- 13
Db 7291 ELLTDLPRPSILSGRAGSVDTTIEGSGVALOSFCHTRSVTTFVILLIVFRIAHPLTAV 7350
QY 14 ----- 13
Db 7351 DDATIGTPIANRNRPELETLVGCFTVNTQCMRISIADDDNFEGLVROVRNVAATAANQDV 7410
QY 14 ----- 13
Db 7411 PFERIVSALVPGSRTSRNPLVQVLMFAVQSVEDYDQVRLEGLSVMPGAEASTRDMERH 7470
QY 14 ----- 13

```

Db 7471 LVPDQKLTGSVLYSSDLFEQGTIONFVDIFQECRLSVLDQPLTPISVLPFSNAISNES 7530
QY 14 ----- 13
Db 7531 LDLEMTSDYPRDRTVYDLFERQQAICPDIAVKDSSQLTYAQDLQDSRVAAMLHER 7590
QY 14 ----- 13
Db 7591 HMPAESLVGVSFSCSETIATFSGIMKANLAYLPDYAPDARLAALDPTVEGERLLLG 7650
QY 14 ----- 13
Db 7651 AGVPOGIIQIRLSTAVIAEALSHATVDTSTIPQPSATSLAVIFTSGTGKPKVWIE 7710
QY 14 ----- 13
Db 7711 HRGIVRLVDTNVNVPESGALPVSHSNLAMDATWEIYTAVLNGGVICIDRDTMD 7770
QY 14 ----- 13
Db 7771 IALNSTFRKENVRAFFTPAFKOCIAETPELVANLEIHTAGDRLDPGDANLAGTKAR 7830
QY 14 ----- 13
Db 7831 GGIFNVLGHTEHTAYSTFFYPVGEETPVNGVPGRGISNGHAYIIDRHOKLVPAGVAGEL 7890
QY 14 ----- 13
Db 7891 ILTGDAVARGYTDSALNKDRVYIDINGKSTWSYRTGDKARPRDQLEFFGHMDQVK 7950
QY 14 ----- 13
Db 7951 IRGVRIEVEVELTLDBKSVLATVYVRRPNDPEMIAFITDADVDQTHKAIYKHL 8010
QY 14 ----- 13
Db 8011 QGILPAYMISHVILDPMPYTDNGKVRDRLALRAQTQKRSTARVPRDEVEAVLC 8070
QY 14 ----- 13
Db 8071 EESYNLEVEVGITDGFDDGSHSLATKLAARLSROLNTRVSXKVDQPIILADIIL 8130
QY 14 ----- 13
Db 8131 RRGSHRDPIRATPYTGPVQSFQAGNLFLEQLNGASWLMFPFIRMGPIQOTALAV 8190
QY 14 ----- 13
Db 8191 ALNALVHRHRLRTTFEDHDGVGVYIQPKSQDLRIIDLSADAVDTAYLAALKRQTTA 8250
QY 14 ----- 13
Db 8251 FDLTSEPGWVSLRLDQDDYIILSYMHIIISDGTVDLROELGOFYSNAIRGOEPLSQ 8310
QY 14 ----- 13
Db 8311 AKSLPIYRDFAWWQROENIKQAKQKXWSQLADSTCEFLTDLPRESIISGEADAV 8370
QY 14 ----- 13
Db 8371 PMVIDGVYQULDFCHTQVTSFVLLAARAHYRLTGLDAGVGTPIANRNPBELG 8430
QY 14 ----- 13
Db 8431 LIGFVNTQCKRAISETETTESISVQOVRTLTTEAFANDVPFQIVSTLLPGSRDTSR 8490
QY 14 ----- 13
Db 8491 PLVQVMEALQSQDLGRIOLEGMTDEALEPLSTRDLQEVHLQOEYKLSGSLIYSTDLE 8550
QY 14 ----- 13
Db 8551 EVETIRGIVDVLEILRRGLEQPKORIMAMPITDGITKLRDQGLTVAKRAPPRESSVID 8610
QY 14 ----- 13
Db 8611 LFRQVAAAPDALAVWSSSTLYADLDQSNKLAHMLCORNAPEPTLVAVPASCCTI 8670
QY 14 ----- 13
Db 8671 VAEGLVKANLAYLPDVNAPARIBAILSAVPGHKLTVQAHGPELGLTMADTELQYD 8730
QY 14 ----- 13
Db 8731 EALASSSGDHEQIHASGPTATSLAYMFTSGTGPKGMIDHRSIIRLVKNSDVATL 8790
QY 14 ----- 13
Db 8791 PFPVRMANVSNLAFDISVOEITALLNGSTLYCLDYLTLIDSKILYNVEVAQVNAFT 8850
QY 14 ----- 13
Db 8851 PVLKQCLGNMFAIISRLSVLFVNGDRDLDAHDAVAASGLIDAVYNAVGPTEMGOSTMY 8910
QY 14 ----- 13
Db 8911 KUDVNEPVPNGVFIGRSTNSGATYMDGNQOLVSPGVGEIVVTGDLARGYTDALDED 8970
QY 14 ----- 13
Db 8971 RFVHTIDGENIKAYRTGDRVRYRPKDEIEFFGRMOQVKIRGRHIEPAVEHALLGH 9030
QY 14 ----- 13
Db 9031 DLVHDAVYLRRKANQPEMIAFITSOEDETIEQHSNKQVQMGHEHPSRYADIKDD 9090
QY 14 ----- 13
Db 9091 TSTFGHDFGWTSMYDQVDIPVNEKEMWDETTSALDNRPPGHILEIAGCTMILSNIG 9150
QY 14 ----- 13
Db 9151 KVDGLQKYVGLDPAPSAIFVNEAVKSLPSLAGKARVLVGTALDIGSLDKNEIPELYVI 9210
QY 14 ----- 13
Db 9211 NSVAQYPTSEYLIRKVAKVEVPSKRVFSGDIRSQUALNRDPLAARAVRALGNASKQ 9270
QY 14 ----- 13
Db 9271 IREKIALESEBELLVPAFFVSLRSQLPNIKHYEVLPKMKATNELSYRYAAVLHS 9330
QY 14 ----- 17
Db 9331 HNEEQLIIDIPTAWVDFAATOKDSQGLRNLQGGRDVMIAGNIPYSKTIYERIM 9390
QY 18 ----- 17
Db 9391 NSLDDHVNSLDGTWISDARSAAICTSFDAPALTOAKEGFRVELSMARORSQNAL 9450
QY 18 ----- 17
Db 9451 DAVFRLATDANCERSRVLVHPTDHQROLRTLNRPLQRAQSRISQVFEALQTLAP 9510
QY 18 ----- 17
Db 9511 AYMIPSRILVLPQMPNTNNGKVKRQLARQAQVAKRRAVSARVAPRNDTEIVCEEYAD 9570
QY 18 ----- 17
Db 9571 ILGTEVGITDNFPDMGSHLATAKLAARLSRLDRTVTYKVEFKPVLADLAASIEGST 9630
QY 18 ----- 17
Db 9631 PHLPIASSVSGPVQSVTAQGRWLFDPFNUNATWYHMSLAMRLGLPLNMDALDVALARL 9690

```

QY 18 ----- 17
Db 9691 EQRHETRTTFEAKODIGVOVHEAGMKRLKVLDSLDKNEKEMAVLENOQMPFLLASE 9750
QY 18 ----- 17
Db 9751 PGWKCHLARIGPTEYILSLVWHMFSDGWSVDILRQELGQFYSAALLGRDPLSOVKPLPI 9810
QY 18 ----- 17
Db 9811 QYRDEAAMQKBAQVAHEROLAYWENQADSTPGELLTDEPRPQFLSGAGVLPVTEG 9870
QY 18 ----- 17
Db 9871 PYEKLKFSKEROVTFESVLLTFAFRATHRLTGAEDATIGTPIANRNPBELEHIIGFFV 9930
QY 18 ----- 17
Db 9931 NTQCMRLLDPTGSFESLVQHVRSVATDAYSNODIPFERIVSALLPGSRDASRSPLIOLM 9990
QY 18 ----- 17
Db 9991 FALHSQDPLNITTEGLEHERLPTSVATRDMFHLFOEPNKLGSITLFADELFOPETIN 10050
QY 18 ----- 17
Db 10051 SVTVFQELRLRGIDQOVQVISTMPITDGLIDLEKLELESSENFPRDYSVYDVFQOY 10110
QY 18 ----- 27
Db 10111 AANPNAPVVDSEMSYTSLSQKSEOIAAMLHAQGLRPESLICVMAPRSPETIVSLGI 10170
QY 28 ----- 27
Db 10171 LKAGVATPLPDVNSPARAIPQILSEVBGKRLVLLGSGIDMPQSDRMDETARIDILTNT 10230
QY 28 ----- 27
Db 10231 KVERSDPMRSPATSLAVITSGTGPRKGVMEHNRILRLVKQSVTSQPLQDLMAH 10290
QY 28 ----- 27
Db 10291 ISNLAFAISWEFTALINGGALICIDYFTLLDSQALRTTEKARVATLFAPALKECL 10350
QY 28 ----- 27
Db 10351 NHAPTLEDLAVLYIGDRDLATDAKIQALVKGYVNAVPTENTVASTIYRLTDESY 10410
QY 28 ----- 27
Db 10411 ANGVPIGNAVSSGAYIMDQORLVPFGMGLVVGSGDLARGYSTNLNADRFVDIVIN 10470
QY 28 ----- 27
Db 10471 DOKARAVTGDRTXRPKDGSIIEFFGRMDQVKIRGHRVEPAVEQAMGNKALIHDAVV 10530
QY 28 ----- 27
Db 10531 VQAVDQETEMIGFVSMASSDFSEGEIEITNOVEMEDHFEFSTAVAGIEALDQATLGRDF 10590
QY 28 ----- 27
Db 10591 TSWTSMYNGNLIDKAEWEMLDITMQSLDKEDARPCAEIGTGWVLFNLPKNDGLESY 10650
QY 28 ----- 27
Db 10651 VGIEPSRSAALEVDKAADFPGLQKQIILVGTAEIDIKLVKDEHDDVVVINSVAQYPPSR 10710
QY 28 ----- 27
Db 10711 SYLVQIASSELHMTSVKTIFFGDMKSMATNDFLVSRLALYTLGDKATQDQIROEVARLEE 10770
QY 28 ----- 27

Db 10771 NEDELLVDPAPFTSLISQWPGVKVHVEILLPKMRTSNELSSVRYAAVLHICRDBGRNRY 10830
QY 28 ----- 27
Db 10831 GRHVHSEENAMIDPSSGMDRHALVOMLDERROAKTVAIGNIPHSNTINERHFTTSIDT 10890
QY 28 ----- 27
Db 10891 EEBGIAQDSLDGSANQSAKAMAARCPCLSTVELVEIGQAAGFVYVSMARQSHGALD 10950
QY 28 ----- 27
Db 10951 VVEHLEDDRVGRVLIINPTDERLPPSTGLTSRPLQIRIONRRFSQIREQLTLLPPYM 11010
QY 28 ----- 27
Db 11011 VPSRIVLERMPLNANSKYDRKELARKARTQITIRPSATRVAPRNDIEAVLCDEFQAVLG 11070
QY 28 ----- 27
Db 11071 VTVGVADNFFELGHSIMATKLAARLSRLDTRVSVKDIENQPILODLADVYQGSAPHE 11130
QY 28 ----- 27
Db 11131 ALPSTPYSGPVQSFQGRUWFLDQNLNASWYHMPLASRLRGLRIBALQSLATIEAR 11190
QY 28 ----- 27
Db 11191 HSLRTFEEODGVPOVQVRAARNQRLIIVSGTEDAVLAALQEDADLTAEPQWR 11250
QY 28 ----- 27
Db 11251 VALLRGPDDHVLSTVMHHIISDGWSVDILRQELGQVSNASSQAPAPLPIQYRDEAIVQK 11310
QY 28 ----- 27
Db 11311 QDSQIAEHOKOLYWKROLVNSKPAELLADFTPRALSGDADVIEIDQVQYMLRSFC 11370
QY 28 ----- 27
Db 11371 RARHVTSEVALLAAPRAHYRLTGAEDATIGSPIANRNPBELEGLICFVNTQCLRIYK 11430
QY 28 ----- 27
Db 11431 SEDTFTLVQARETATEAQDNQDVPFERIVSSVASSRDTSRNPLVOVMEAVHSQHDIG 11490
QY 28 ----- 27
Db 11491 NTRLGVBGKPVSMASSTFDEMLFEDQMLGCVNVSFKDLFESETIRSVAAVFOETL 11550
QY 28 ----- 27
Db 11551 RRGIANPHANLATPLTDLPLSLRSLCLOVNPDPVPRDASVIDVREQVASTIPKSAVID 11610
QY 28 ----- 27
Db 11611 ASSQLYTELDERSQOLATWLRQVTVPEELVGLAFRSCETIIAFLGIKANLAYPLUD 11670
QY 28 ----- 27
Db 11671 VNAPAGRIETILSLDGNRLILGSDTOAVKLANSVRFTRISDALVESGSPTEELSTR 11730
QY 28 ----- 27
Db 11731 PTAQSLAYVWFISGSTGVKGVVVEHRTIRLVKNSNVYAKQPAALAHLSNIAFDASS 11790
QY 28 ----- 27
Db 11791 WEIYAPLNGVVCIDYTTTIDIKALEAVFKOHIRGAMLPALLKQCLVSAPTMISL 11850
QY 28 ----- 27

Db 11851 EILFAAGDRLLSSDAILARRAVSGVYNAYGPTENTVLTSTIHNIGENAEFSGVPIGNAY 11910
 QY 28 ----- 27
 Db 11911 SNSGAFVMDQNOQLVAGVIGELVVTGDLARGYTDSKLRVDRFTYITLGDGNVRAVARTG 11970
 QY 28 ----- 27
 Db 11971 DRYNRPRKQIEFFGMDQIKIRGRHIEPAVEQALANDPAISDAVITQLTDEEREPE 12030
 QY 28 ----- 27
 Db 12031 LVAFESLKGANANGNGVNGVSDQEKIDGQHALMKNKIRHMLQALLPTMIPSRITIV 12090
 QY 28 ----- 27
 Db 12091 DOLPVNANGKIDRNELAVRAQATPRTSSVSTYVAPRNDIETICKEPADILSVRGITDN 12150
 QY 28 ----- 27
 Db 12151 FFDLGGHSLIATKLAARLSRRLDTRVSVRDVDTPVVGQLAASIQGSGPPEHAIPALSHS 12210
 QY 28 ----- 27
 Db 12211 GPVQGSFAGQRIWFLDRFNINAAWYIMPEGVRLRGPLVDALQTLRALERHLLRTTF 12270
 QY 28 ----- 27
 Db 12271 EEDGVGMQIVHSPRMDICVVDISGANEDLAKLKEEQAPFNLSTEVAVRVALFKAGEN 12330
 QY 28 ----- 27
 Db 12331 HHLLSTYMHIIISDGSVDIFQOELAQFSAVARGHDPISQVPRPIHYRDFAVWQRODK 12390
 QY 28 ----- 27
 Db 12391 QVAVHESQIQWTEQLADSTPAELSDENRPEVLSEAGTVPIVIDEYVEKLSLFCRMH 12450
 QY 28 ----- 27
 Db 12451 QVTSFVLLAFAVAVHYRLTGADATIGPIANRNPEDLDLIGFVNTQCMRIALEHD 12510
 QY 28 ----- 27
 Db 12511 NPLSVRRVASTAASAFENODVFERLVSAALLFGSDASRNPLVOLMFVHSGRNIGKIQ 12570
 QY 28 ----- 27
 Db 12571 LRLGEGBPPTATATREDFVEFHLFEQDKGLAGNVFAADLFEATIRSVVEVHEILRGG 12630
 QY 28 ----- 27
 Db 12631 LDQPDIAISTMPLVDGLAALNSHMLPAVEDIEPPFATEASVVDVQFQVVPDADALAVTD 12690
 QY 28 ----- 27
 Db 12691 TSTKLTVAELDQSDHYAAMLKQKLPASTIYVVLAPRSSETIVACIGILKANLAVLPMD 12750
 QY 28 ----- 27
 Db 12751 SNVPEARQAILSEIPEGEKVFLLGACVPIPDNKTADVRAVFTSIDIVASKTDKSYSPGTRP 12810
 QY 28 ----- 27
 Db 12811 SASSLAVYFTSGSTGRPGVMEHRRGIVSLVKQNASRIPQSLRNAHVSNLAFDAVWEI 12870
 QY 28 ----- 27
 Db 12871 FTTLLNGTLFCISYFTVLDKRALSAFSDHRINTILPLPALIKQCLADAPSVLSLESL 12930
 QY 28 ----- 27
 Db 12931 YTGDRLDGADATKVKDLVKGRAVNAAGPTENSVMSTIYTIHETFANGVPIGTSLSGPKS 12990

QY 28 ----- VGRF 31
 Db 12991 KAVIMDQDQLYPAGVMEGLVYAGDGLARGYTDPSLNTGRF 13031

RESULT 3

T36248

CDA peptidase synthetase I - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000

C:Accession: T36248

R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, submitted to the EMBL Data Library, March 1999

A:Reference number: 221602

A:Accession: T36248

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-7463 <SAU>

A:Cross-references: EMBL:AL035640; PIDN:CA818518.1; GSPDB:GN00070; SCOEDB:SC563.03C

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: cdaS1, SCOEDB:SC563.03C

C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:516-1074/Domain: acetate-CoA ligase homology #status atypical <ACLI1>

F:1090-1158/Domain: acyl carrier protein homology <ACPI>

F:1715-2184/Domain: acetate-CoA ligase homology <ACLI2>

F:2200-2268/Domain: acyl carrier protein homology <ACPI2>

F:2804-3249/Domain: acetate-CoA ligase homology <ACPI3>

F:3265-3332/Domain: acyl carrier protein homology <ACPI3>

F:4323-4746/Domain: acetate-CoA ligase homology <ACPI4>

F:4762-4830/Domain: acyl carrier protein homology <ACPI4>

F:5363-5786/Domain: acetate-CoA ligase homology <ACPI5>

F:5802-5870/Domain: acyl carrier protein homology <ACPI5>

F:6401-6868/Domain: acetate-CoA ligase homology <ACPI6>

F:6884-6951/Domain: acyl carrier protein homology <ACPI6>

F:1122-2232, 3297, 4794, 5834, 6916/Binding site: phosphopantetheine (Ser) (covalent) #st

Query Match 42.9%; Score 67; DB 2; Length 7463;

Best Local Similarity 0.5%; Pred. No. 3, 5e+03;

Matches 17; Conservative 2; Mismatches 8; Indels 3127; Gaps 3;

QY 4 HXSMEXRTP ----- 13
 Db 3537 HXSMEXRTPHTLRDPARDTHATAGETITLALDPTTALLTWPGCHATVNDVETSTF 3596
 QY 14 ----- 13
 Db 3597 ALAVAGRRGRGADADAVLVLDSEHGRHEBAVPGVELSRTAGWFTSMYPVRLAPAGAS 3656
 QY 14 ----- 13
 Db 3657 GDGSAIRALKAVKEQKRTVPGDGLGYGLLRHLNPTRAALALPLPEGFNYLGRIGQEG 3716
 QY 14 ----- 16
 Db 3717 TDEAPWITIEGDVAGIDGAMPFAHPVDVNAVAFETADSTRARWTSRTALEPEDTQRL 3776
 QY 17 ----- 16
 Db 3777 ADTWFRLLRLVEAROPAGAGLTPSDIAPHALAOEDTEDELTHTVPGIADILPLAPLQEG 3836
 QY 17 ----- 16
 Db 3837 FLFLNLVDENARDVYVGGIADLDGSPDGTMRRAAGALLRRHANLAPRGQTATGTWQ 3896
 QY 17 ----- 16
 Db 3897 VVPALEPDWRECDLTDRADEARDAEAGRLAAGDRERREDLTSPLMRTAIRLSADRY 3956
 QY 17 ----- 16

```

Db 3957 RLVMTNHHILLDGMSPMLMQLMELTLYVSGGDVSLPVPARYRDHLAMLGARDARDA 4016
QY 17 ----- 16
Db 4017 WRRSLSGIDEATLLAPDAGFAEAPLIGPGLDRDATALSAMARGVMTNTVOGANA 4076
QY 17 ----- 16
Db 4077 LALAQATGRDDVYFGATVSGRPPELPGVESMIGLFINTLFPYARLDQAEPLDLPFRLO 4136
QY 17 ----- 16
Db 4137 EQARLDHQPGLADIQHMAGHGLFDTAMVFOVPVADTTSROLGLKVGAYDAVEST 4196
QY 17 ----- 16
Db 4197 DPAVNLVAHTRDALRLRLDYRADACAGDLVRSIADRMLEALVYDSDRPVAILDTLD 4256
QY 17 ----- 16
Db 4257 PAVREYLVEMNGAPQLPGTPLHLLISEQARLTPDAVAVVCGTSLITYAELDGGANOLA 4316
QY 17 ----- 16
Db 4317 RHLLEGIGAEDEVAIALAKSLDAVISMLAVLKTGAAYLPIDDPYAEKITVMDARPA 4376
QY 17 ----- 16
Db 4377 LTLTEPVVERTYGHSTAVTDERNSPWSARHAAYMTTSSGTRPKGVYIEHHALATY 4436
QY 17 ----- 16
Db 4437 LHRANTYTMGTGYVLSPLAFDLITALTMTPLTSGGVHLTSLSESTQPSLIKATPS 4496
QY 17 ----- 16
Db 4497 HLPLTLPLTPASPHTLLIGGALHTDHLVTWTOHPGOIINAYGPTESTVNTDHHV 4556
QY 17 ----- 16
Db 4557 GEDTPDGPVPIGRFANTQYVYVLSALRPVAPVGTGELYLAGDQLANGYLGRRALTAERF 4616
QY 17 ----- 16
Db 4617 TANPSSTPGARMTGTDLAMHMHGHLTYDGRADHOKLGRHIEGEIEATLTAOTGI 4676
QY 17 ----- 16
Db 4677 TQATVOLREDOPGDORLVAVLVNDSTEYDEKTVRDALTSALPDYVPSALVTLLALPLT 4736
QY 17 ----- 16
Db 4737 PNGKIDRTALPAPAYASASTAGRAPRTPREVICTLFAEYVGVDTVTIDNFFDLGSHSL 4796
QY 17 ----- 16
Db 4797 ATRVSRTRTALGVELSIROLFETPTVAGLAELADAGTWTALTARPERIRPLSYAOQ 4856
QY 17 ----- 16
Db 4857 RLWFLHOLEGPSANTVNLRLIGCALDVADALRAISDVVARHESLFTVTEDERGAYOI 4916
QY 17 ----- 16
Db 4917 VLPVEASTPTTYVDVAEEELGDRLEAVGHCFLDQELPARISLFFVSEREHVLILLIH 4976
QY 17 ----- 16
Db 4977 HIASDANSRAPLADLTAAAYAAVRSAPMWAPLTVQADYALMQOEILGDDTDADSLAG 5036
QY 17 ----- 16
Db 5037 ROLAYWMOQLAGLPEQDLPTDRPRPAVAGYSGDRVPFTVPELHTRLTELARATNTSAF 5096
QY 17 ----- 16

QY 17 ----- 16
Db 5097 MYIOAAVAVLLTRLGAGEDIPIGTPVAGRTDDAADLIGLFINTLVLERTDTSQDPTERRL 5156
QY 17 ----- 16
Db 5157 LDRVRDIDLAAYAHQDLPERLVEALNPARTLSHPLFQVLLTFNNTDHGCALKDISELP 5216
QY 17 ----- 16
Db 5217 GLTVALREVORTSKFDLSFGFAESFDTSRROGLEALDSTELDRSQAOLADRFLLR 5276
QY 17 ----- 16
Db 5277 VLEANTAPDRPIGAVELMDPAERERVLVEMNGAPQLPGTPLHLLISEQARLTPDAVAV 5336
QY 17 ----- 16
Db 5337 VCDGTTLYAELDRRANOLARHLLEGIGAEDEVAIALAKSLDAVISMLAVLKTGAAYLP 5396
QY 17 ----- 16
Db 5397 IDDPYAEKITVMDAQPALTLAPIPASYSRPTSEITDVERRSPWSARHAAYMTT 5456
QY 17 ----- 16
Db 5457 SSGTRPKGVYIEHHALATYLRANNTYTMGTGYVLSPLAFDLITALTMTPLTAGTV 5516
QY 17 ----- 16
Db 5517 HLTSLEAEVOPSLIKAPSHLPLTLTPASPHTLLIGGALHTDHLATWTOHPGA 5576
QY 17 ----- 16
Db 5577 QLIINAGPTESTVNTDHHVSEDPDGPVPIGRFANTQYVYVLSALRPVAPVGTGELYL 5636
QY 17 ----- 16
Db 5637 AGEOLARGYLRPALTAERFANPHSSTPGAMRYRTGDLAMHNDHGHLTYDGRADHOKL 5696
QY 17 ----- 16
Db 5697 RGRHIEGEIEFTTLAQGIGIQTVOLREDPGDORLVAVLVNDSTEYDEPTLRLDALAS 5756
QY 17 ----- 16
Db 5757 ALPDYWRPSAYVTLDALPLTPNGKIDRTALPAPAYASASTGTGRTPTPREBELICTLFAEVL 5816
QY 17 ----- 16
Db 5817 GVDLVYIDNFFDLGSHSLATRLVSRARTALGVELSVROFEFPTIAGSGAEDRAGRA 5876
QY 17 ----- 16
Db 5877 RAALTAARPERIRPLSYAOQLWFLHOLEGPSANTVNLPTLRLGTLDTLDALQSALNDLL 5936
QY 17 ----- 16
Db 5937 ARHESLRTTYTDDGSPROYIHAMEBGMLPLGVYDTGSELDAMLSAGVHNHAFDLTAGIP 5996
QY 17 ----- 19
Db 5997 VRATLFRISEQEHVLLLIHIIATDMSRTPLGHDIAAAYSARCAGDVPAMEPLRQYAD 6056
QY 20 ----- 19
Db 6057 YALMOREVLGDEGDADAPAGROLAYWTRQOLADLPEQLDPLDRPRPAVASDGDGRVAFSL 6116
QY 20 ----- 19
Db 6117 DADLYVRLELARATHSSITFMVVOAALAVLLTRLGAGEDIPIGTPVAGRTDADENLVGF 6176

```

```

Qy 20 ----- 19
Db 6177 FVNTVLNNDTSNGPTEFELLETRRDLAAYAHQDLPFEFLVYALNPARTLAHHPLEQV 6236
Qy 20 ----- 19
Db 6237 MLISTETDASIALPGLNVAERSLCAKAKVDLAPALAEVNDGEGRTGLTGLDIFR 6296
Qy 20 ----- 19
Db 6297 TDLFDRSTARSLVERFYRTLEAVVADPGVRLSRVPLVLSGERSRLDRGTGPILEGIDAT 6356
Qy 20 ----- 19
Db 6357 LPELFAQALTPGAPALVRGCTVSAYAEIDLRTNRLARLLRQGVNPGTVPVWLMERSP 6416
Qy 20 ----- 19
Db 6417 AHVAVTALAKAGAGAYVPLHDTYLYLDMRRHVADTAATLLITDRAEARAGOLGARVMV 6476
Qy 20 ----- 19
Db 6477 DEFGAAPSSEADAPGCTGTGTGTSRSGYVDAPVEYGLRPODLAYVMTSGTGVPKGV 6536
Qy 20 ----- 19
Db 6537 AVTHRGVVDLVDRHCHMPRGVHERVLLHAPAFDVSCEMMVPLVSGTGVVAPGHLDP 6596
Qy 20 ----- 19
Db 6597 AITDLTAHDTAHLTACGFRVVAEAPCEFCAGREVLTGGOVSPAAVAVLAHHPRI 6656
Qy 20 ----- YXKRGIRPVGR 30
Db 6657 VLRLHVGPTETTLCTVQHEVTAPEARGLSPVGR 6690

RESULT 4
I38344
titin, cardiac muscle [validated] - human
N:Alternate names: connectin
N:Contents: serine/threonine-specific protein kinase (EC 2.7.1.-)
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 15-Sep-2000
A:Accession: I38344; I38345; S20898; S20899; S63665; S37393
R:Label: S.; Kometer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MID:96026330; PMID:7569978
A:Accession: I38344
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
A:Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425
R:Musco, G.; Tzitzizios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: Identification of an alpha-helix
A:Reference number: I38345; MID:95119041; PMID:7819249
A:Accession: I38345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:g602579; PID:CAA58243.1; PID:g602580
R:Label: S.; Gautel, M.; Lakay, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MID:92258380; PMID:1582406
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 13597-14200; '1', 14203-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:g37192; PID:CAA45939.1; PID:g37193

```

```

A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:g37190; PID:CAA45940.1; PID:g37191
A:Accession: S20899
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 2248
A:Cross-references: EMBL:X64697; NID:g37190; PID:CAA45938.1; PID:g37195
R:Kometer, B.; Olivieri, N.; Wilt, C.C.; Herrmann, B.G.; Labelt, S.
J. Mol. Biol. 256, 556-563, 1996
A:Title: Genomic organization of M line titin and its tissue-specific expression in t
A:Reference number: S63665; MID:96177761; PMID:8604138
A:Accession: S63665
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 26729-26825 <KOL>
A:Cross-references: EMBL:X92412; NID:g1236761
R:Gautel, M.; Leonard, K.; Labelt, S.
EMBO J. 12, 3827-3834, 1993
A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in different
A:Reference number: S37393; MID:94008990; PMID:8404852
A:Accession: S37393
A:Molecule type: mRNA
A:Residues: 26831-26926 <GAU>
R:Improta, S.; Poltun, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A66736; PDB:1IT7
A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R:Piñol, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A66201; PDB:1NCR
A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q32
C:Function:
A:Description: structural protein forming filaments in striated muscle
C:Superfamily: titin, fibronectin type III repeat homology; immunoglobulin homology;
C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; g1
structural protein
F:24752-25008/Domains: protein kinase homology <KIN>
F:84,177,905,2276,2378,2459,2481,2563,2763,2896,3088,3179,3384,3432,3628,3772,40
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,1354
tatus predicted
F:16780,16976,17579,17602,17667,17681,17685,17899,18121,18188,18209,18336,18670,18680
,21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,
F:26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experiment
Query Match 42.3%; Score 66; DB 1; Length 26926;
Best Local Similarity 0.1%; Pred. No. 4,8e+04;
Matches 22; Conservative 3; Mismatches 6; Indels 24391; Gaps 7;
Qy 1 SXXHX----- 5
Db 2459 NRTFASDEGPKYLIYGRVETNCNLSEKIKITRGLDRLCTETQNVVPEVLSHGIDVL 2518
Qy 6 ----- 5
Db 2519 WNFKDEKIPSSKYIEAHGKIYKLTILNMKDRGKYFAGENMTSGKLTVAAGALSK 2578
Qy 6 ----- 5
Db 2579 PLTDQVAESQAEVPECEVANDPSKGEWLROCKHLPPLNNIRSESDGKRRLLIAATKLD 2638
Qy 6 ----- 5
Db 2639 DIGEYTKVAASKTSAKLKEVAVKIKTKILVTETQDAVFVELTHPNVKGQWIKNG 2698
Qy 6 ----- 5

```

Db 2699 VLESNEKATISVKGITYSIRKNCALVDESVEGFRIGRLGASARLHVETVKKIKKPKDV 2758
QY 6 ----- 5
Db 2759 TALENATVAFEVSVSHDTPVFKWFKSVKSPDKHRLVSEKVKHKLMLONISPSDAGEY 2818
QY 6 ----- 5
Db 2819 TAVVGOLECKAKLFEVETLHITKMKNIEVPETKASFCEVSHFVPSMMLKNGVELEMS 2878
QY 6 ----- 5
Db 2879 EKFKIVOGKILHOLIMNTSTEDSAEYTVCGNDVSAITLVPIMITSMUKDINAEKD 2938
QY 6 ----- 5
Db 2939 TITEEVTVNEGISTYKMLKNGVEITKSDCKOMRTKLLTHSLNINNHFGDAADYTVAGK 2998
QY 6 ----- 5
Db 2999 ATSTATLVEARHIEFRKHIDIVLEKKRAMFCEVSEPDITVQMKDDELQITDRK 3058
QY 6 ----- 5
Db 3059 IQEKRYVRLIPSTRMSDAGKTYVAGNVSTAKLVEGHDVIRISIKKEVOYIEKORA 3118
QY 6 ----- 5
Db 3119 VVEFEVNEDDDAHMYKKGIEINFOVERHKYVYERRIHMFISETRSDAGEYTVAGR 3178
QY 6 ----- 5
Db 3179 NRSVTLVNAPEPVOVLQELPVTVOGKPARFCAMISGRPKISWYKEOQLSTGFK 3238
QY 6 ----- 5
Db 3239 CKFLHGOGEYTLILLIAPEDANAYYTCBAKNDYVATTSASLSVEVPEVSPDOEMEYTP 3298
QY 6 ----- 5
Db 3299 PAITPLQDVTYSEGOPARFCQVRSVGTDLKVSWSKDKIKRPSFRMTQEDTYOLEIA 3358
QY 6 ----- 5
Db 3359 EAYPEDEGTYTFVANNAGVYSTANLSLEAPESILHERIOEITEMEKESFSLAE 3418
QY 6 ----- 5
Db 3419 EGLHSAELQLSKINETLELLESBPYPTPKDSEKGTGPIFIKEVSNADISMGDVATLSV 3478
QY 6 ----- 5
Db 3479 TVIGIPKRIQWFFNGVLLTPSADYKFVFGDDHSLILFTKLEDEGEYTCMASNDYGT 3538
QY 6 ----- 5
Db 3539 ICSAVLAKINSKGEHKDTESENAKSLKLGCPPHFLKELPIRCAGLPAIFEXTV 3598
QY 6 ----- 5
Db 3599 VGEPAPTYWREKQLOCTSVYTTIIHNPNGSGPIYVNDPQREDSGLYICAAEMMLGEST 3658
QY 6 ----- 5
Db 3659 CAEELLVLEPDMDTCPKAKSTPEADEPQPPLKGPAAVELDSEQIATFVKDTLK 3718
QY 6 ----- 5
Db 3719 AALITEENOQSYEHIAKANELSSQLPLGAOELQSLQEDKLTPESTREFLCINGSIHQ 3778
QY 6 ----- 5
Db 3779 PLKEPSPNLQLOIVOSQKTFPSKEGILMPEEPETOAVLSDTERKIPPSAMSIQINSIYVER 3838

QY 6 ----- 5
Db 3839 IKTLLAPEGNTPOSSIEPPMHSTYLSVAEVLSTKEKTVSDTNRQVTVLOKQOASAL 3898
QY 6 ----- 5
Db 3899 ILSQSLAEGHVESLOSPDVMIQOVNVEPLVPSEHCTEGCKILLIESANPLENAGDSAVR 3958
QY 6 ----- 5
Db 3959 IEEGSLRPLALEEKQVLLKEEHSNDVMPDQIIESKREPAIKVOEVQGRDLISKE 4018
QY 6 ----- 5
Db 4019 SILSGIPEORLNTKIQICRALQAAVASRQGLFSEMLNIEKEVEAVNITQEPRHIMC 4078
QY 6 ----- 5
Db 4079 MYLVSATSVTEEVYIIIEDVDPOMANLKMELDALCALIYEIDILTAEGPRIQOGAKT 4138
QY 6 ----- 5
Db 4139 SLOEKDSFGSOKVEPITEPEVESKYLISTEBSYTFNVQSVKYLDTATPVTKVASAVV 4198
QY 6 ----- 5
Db 4199 SDEKQESLKPSEKESSESSEGTAEVATVKIQEAGGLIKEDGPHHTPLVDTVSEGD 4258
QY 6 ----- 5
Db 4259 IYHLLTSTNAKEVNVYFENKLVPSDEKFKLODONTYTLVIDKVNTEHDGGEYCEALN 4318
QY 6 ----- 5
Db 4319 DSGKTATSAKLIVYKRAAVYIKRIEPLVALGHLAKFCEIQSAPNVAFQWFKAGRELY 4378
QY 6 ----- 5
Db 4379 ESDKCSIRSKYISLSLEILFTQVVDGEYTCRASNEGVSCTANLTVVPGSEKRVKL 4438
QY 6 ----- 5
Db 4439 LPERKPEKEVVLKSLKRRBEEPKYEPKLEKVKPAVPEPPKAVEVEVPYVT 4498
QY 6 ----- 5
Db 4499 KREKTIPEPTKYVEIKPAIPLPAPBPKPBAEVKTIKPPVEPEPTPIAAVTVPVGK 4558
QY 6 ----- 5
Db 4559 KAEAKAPKEBAKPKGPIKGVPKTPPIEAERRKLRPGSGEKPDEAPFTYOLKAVPL 4618
QY 6 ----- 5
Db 4619 KFYKEIKDIIILTESEVGSALFECLVSPSTAITTMKGSNIRBSPKRFIADGDKRL 4678
QY 6 ----- 5
Db 4679 HIIDVQSDAGEYTCVLRGNKEKSTAKLVVEELPVREVKTLLEEYVVKQOPLYLSC 4738
QY 6 ----- 5
Db 4739 LNKERDVWRKDKIYVERGRIVPGVIGLMRALTINDADDTAGTYTVENANNLECS 4798
QY 6 ----- 5
Db 4799 SCVAVVEVIRDMVLKPIRQOHVYKPGTALFACDIADPTNIMFKGYDEIPAPNDKTEI 4858
QY 6 ----- 5
Db 4859 LRDGNHLYLKIRKNAMEPDIAEYAVEIEGKRYPAKTLFGEVEVELLKPIEDVTIYKESAS 4918

QY 6 ----- 5
Db 4919 FDAEISADIPGOMKLGELLRSPCEIKAEGBKREPLTHKVKLDQAGEVLYOALNALT 4978
QY 6 ----- 5
Db 4979 TALLTVEIELDRAVPLKDYTPERROARPECVLTREANNTWSKDPDIKSSDKFDIAD 5038
QY 6 ----- 5
Db 5039 GKKHLLVINDSFDDGEGVTAEEVGKKTSAFLPVTGIRLKFMSPLEDDQYKEGETANTFVC 5098
QY 6 ----- 5
Db 5099 ELSHEKMHVYFKNDAKLHSTRVLISSEGTHTLMEKEVTLDDISOIKAQVELSSTAQ 5158
QY 6 ----- 5
Db 5159 LKYLEADPYFTVKLHDKTAVEKDEITLKCEVSKDVPYKFKDGEIYVSPKYSIKADGLR 5218
QY 6 ----- 5
Db 5219 RILKIKKADLKDRGEVYCDGTDKTKANTVEARLIEVEKPLYGEVYGETAHFEIELS 5278
QY 6 ----- 5
Db 5279 EPDVHGOMKLGOPLTASPCETIIEGKKHLLHNCQLGMTGEVSFOANAKSAANLKY 5338
QY 6 ----- 5
Db 5339 KELPLFITLSDYKVEKDEAKFECEVSRPEKTFRMLKTOEITGDDRELKDGTHS 5398
QY 6 ----- 5
Db 5399 MYKSAFEDAEAKYMEAEDEKHTSGKLLIEGIRLKEFLPLKDYTAKEKESAVTVELSHD 5458
QY 6 ----- 5
Db 5459 NIKRVKFKNDORLHTTRSVSMODEGKTHSTFEKDLSDTSTQIRVAMGMSSEAKLTVLE 5518
QY 6 ----- 5
Db 5519 GDPYFTKLDYTVKEDEVILIOCEISKADAPYKFKDGEIKRPNNAVTKIDGKRMLI 5578
QY 6 ----- 5
Db 5579 LKRALKSDIGQYTCDCGTDKSGKLDIEDREIKLVPLHSVEVMEETAFETEISEDDI 5638
QY 6 ----- 5
Db 5639 HANMKLKGALLQTPDCEIKEBGKIHSLVLAHNCRLDQGTGVDFOANAKSSAHLRVKPNV 5698
QY 6 ----- 5
Db 5699 IGLRLPLKDYTAGETATFDCELSEDIPEVYTKGKLEPSDKVVPSEKGVHTLILR 5758
QY 6 ----- 5
Db 5759 DVKLEAGEVOLTAQDKFTHANLFVEKPPVEFTKPLEDQVEBGATVALECEVSRENAKV 5818
QY 6 ----- 5
Db 5819 KWEKNGEILKSKKYEIVADGRVRLVIHDCTEPEDIKITYTCAKDKRTSCNLTNPVPHVE 5878
QY 6 ----- 5
Db 5879 FLRPLDLQVREKEMAFECELSRENAKVWFKDGAELKGGKYDILSKGAVRILVINC 5938
QY 6 ----- 5
Db 5939 LLDDEAEYSCEVTRTSGMLTVLEBAVFTKLANIEVSEDTIKLVCEVSKPGAELVIM 5998
QY 6 ----- 5

Db 5999 YKDEEIIETGRYEILTEGRKRILVIONAHLEDAGNMCRLPSSRTDGKRVHLEAEFI 6058
QY 6 ----- 5
Db 6059 SKPQNLLEIGEKAFVCSISKESEFPVOMKRDKTLESGDYDIADGKKRVLVVKDRTL 6118
QY 6 ----- 5
Db 6119 QDMGTYVWVGASARAHAHLTVIEKLRIVVPLKDFRVEQOEVENCEVNTGSAKAKWPRN 6178
QY 6 ----- 5
Db 6179 EEAIFDSSKYILLQKDLVYTLRIRDAHLDDQANVNSLTNHERGENKSAANLIVEEDLR 6238
QY 6 ----- 5
Db 6239 IVEPLKDIETMEKSVTWCQKVNRLNLTLMTKNGEVPEDNRVSYRVDKYKMLTIKDC 6298
QY 6 ----- 5
Db 6299 GFPDGEYIVTAGODKSAVELLIIEAPTEFVHLEDQVTFEDDAVFSCQLSREKANKW 6358
QY 6 ----- 5
Db 6359 YNNGREIKGKKYKPEKDSIHLRIKDCRLDCEYACGEDRKRARLFEVEIPEEI 6418
QY 6 ----- 5
Db 6419 RPPDILEAPGADVFLAELNKDKVEVOMLRNNMVVVGDKHOMSEKIHRLQICDIKP 6478
QY 6 ----- 5
Db 6479 RDGEYRIANDKEARAKLELAAPKIKTADQDLVVDYKPLTWVVPYDAPKAAEMFK 6538
QY 6 ----- 5
Db 6539 ENEPLSTKITDTTAAQTSFRILEARKGDKGRYKIVLQKHGKABGFINLKYIDVGPVRN 6598
QY 6 ----- 5
Db 6599 LEVTEFDGEVSLAMEEPLTDGSKIIGYVVERBRIKKTAVLATDRAESCFTVTGLQK 6658
QY 6 ----- 5
Db 6659 GGEYELFRYSARNRVGTGPVETDNPVEARSKYDVPGLNVTITDYNRFGVSLTWEPPE 6718
QY 6 ----- 5
Db 6719 YDGAELITNYIELDKTSIRMDTAMTVRAEDLSATVDVEGQESRVRARNGRIGVK 6778
QY 6 ----- 5
Db 6779 PSAATPFVKVADPIERPSPVNLJSSDOTOSSVOLKMEPPLKDGSPILGIIIERCEBGK 6838
QY 6 ----- 5
Db 6839 DMWIRCNKLVBELTYKVTGLEKGNKYLRYSAENKAGVSPSEILGPLTADDAFVEPTM 6898
QY 6 ----- 5
Db 6899 DLSAFKGLLEVIVPNPITILVPSIGYPRPTATWCPGKRVLETGDRVKMKTLSAVALVIS 6958
QY 6 ----- 5
Db 6959 PSEKSDGIYTLKLENRVKTSIGEIDVNVIAKSPAKELKFGDITKDSVHLTWEPDDG 7018
QY 6 ----- 5
Db 7019 GSPILGYVVERKREVSRTKWTVMDFVTDLEFTVPLVOGKEVLFKVCARNNCGEPBAYV 7078
QY 6 ----- 5

Db 7079 DEPVNKSIPATVDPDPENKWKDRTANSIFLTWDDPKNDGSRIGYIVERCPRGSDKW 7138
 QY 6 ----- 5
 Db 7139 ACGEPAETKMEVTOGLEBGKWAVRVKTLNROGASKPSRPEIOAVTOEAPETFLDYK 7198
 QY 6 ----- 5
 Db 7199 LLAGLVKAGTKIELPAVTVGKPEPKITWTKADMILKODRITIBNVPKKSTVITVDSK 7258
 QY 6 ----- 5
 Db 7259 SPTGTIIEAVNCGRAVAVEVNLDRGPPAADIITDVINESCLLTWNPBRDDGSKI 7318
 QY 6 ----- 5
 Db 7319 TNYVERRATDSEVWHKLSSTVKDINFKATKLIPIKXEYIFRYAENMTGAGEPVQASPI 7378
 QY 6 ----- 5
 Db 7379 AKYQDPGPPTRLBPSDITKDAVTLTWCEPDDGSPITGYVERLDPDIDKWCNKM 7438
 QY 6 ----- 5
 Db 7439 PVKDTTYRVKGLTNKKKTRFVLAENLAGPKPSKSTEPILIKPIDPPWPGKFTVKDY 7498
 QY 6 ----- 5
 Db 7499 GKTSVRLMNTKEHDGAKIESYIEMLKGTDEWVRVAEGVPTQHLPLGIMEQESF 7558
 QY 6 ----- 5
 Db 7559 RYRAVNKAGESEPSBPDLVLCREKLYPPSPRMLEVINITKNTADLKWTVEKDGSP 7618
 QY 6 ----- 5
 Db 7619 TNYIYEKRDVRRKQVDTYKDTKCTVPTLTSGLYFVRYAENAIQSDYTEIEDSV 7678
 QY 6 ----- 5
 Db 7679 LAKDTFTTGPYPYALAVDVTKRHVDLMEPPKNDGGRPIQRYIEKKERLGTWKVAKG 7738
 QY 6 ----- 5
 Db 7739 TAGPDCNFRVTDVIEGTEVOFVRAENAGVGHSEPTIELSIEDPTSPSPDLHTVD 7798
 QY 6 ----- 5
 Db 7799 ACRKHIALAMKPEPKNGSPITGIYHEMCPVGTETKMMRVNSRPIDLKFVYEGVYDKE 7858
 QY 6 ----- 5
 Db 7859 YVLRAVNAIGVSEPSISENVNAKDPDCKPTIDLETHDIIIVIEGKLSIPVPPRAV 7918
 QY 6 ----- 5
 Db 7919 PTVSHKDKAEVKASDRILMKNDHISAHLEVPKSVRADAGIYITLLENKLSATASINVK 7978
 QY 6 ----- 5
 Db 7979 VTGLGPKCDIASDITSSCKLTWEPPEPDGTPILHYLERREAGRTYIIPWMSGNK 8038
 QY 6 ----- 5
 Db 8039 LSWTKDLIPNGEYFPRVAVKVGGEYIELKNPIAODPKOPDPVDVENVHPTAA 8098
 QY 6 ----- 5
 Db 8099 KTIWKPPLYDGSKIMGYIIEKIAKGERMRCNEHLVILITYAKLEGEKQOFVR 8158
 QY 6 ----- 5
 Db 8159 AENAGISEPSRATPTKAVDPIDAPKYLRTSLEVRGDEIADASISSPYPTTWIK 8218

QY 6 ----- 5
 Db 8219 DENVIEEIKKRAAPLYRRRKGEVQEEPPVLPLQORLSIDSKKGESOLARVDSLRD 8278
 QY 6 ----- 5
 Db 8279 HGLYMKVENDHGIAKAPCTSVLDPGPPINFVFEDIRKTSVLCWEPPLDDGSEIIN 8338
 QY 6 ----- 5
 Db 8339 YTLEKKDKTPDSEWIVYVSTLRHCKYVTKLIEGKEYLERVAENRFGGPPCVSKPLV 8398
 QY 6 ----- 5
 Db 8399 AKDPGPPADPKPIVEDVTSNGLVKNWPKNDGSPILGYWLEKREVNSTHMSRVNKL 8458
 QY 6 ----- 5
 Db 8459 LNALKANVDGLBGLATYFVRVCAENNAGPKFSPSPDPTAHDPISPGPPIPVYDTS 8518
 QY 6 ----- 5
 Db 8519 TTIELEWEPAPNGGEIYGVYDKOLVGTNKSCTEKMIKYQYTVKEIRBGADYKL 8578
 QY 6 ----- 5
 Db 8579 VSAVNAAGEBPFGHTOPVTAEPQEPRAVELDVYKGIQIMAGKTLRIPAVTVGRPVPT 8638
 QY 6 ----- 5
 Db 8639 KWTKEGELDKDRVIDNVYNGTSELIKDALKRDHGRVYTATNSCGSFAARVEVD 8698
 QY 6 ----- 5
 Db 8699 VPGEVLDLKPVYVNRKMKCLNMSDPEDDGSSEITGIIIEKDAKMTWOPLETESKCD 8758
 QY 6 ----- 5
 Db 8759 ITGLBGEYKFRVIAKNGKCGPVEIGPILAVDLPPTSPERLTYYERQSTITLWD 8818
 QY 6 ----- 5
 Db 8819 KEPSNGSGSPIOGYIIEKRRHDKPDERVNRKLPPTSPLENIDHOMEFKAVANEI 8878
 QY 6 ----- 5
 Db 8879 GESEPSLPLNVYIODEVPTIKRLSVRGDTIKVAGEPVHIDVYGLPMKIEWSKN 8938
 QY 6 ----- 5
 Db 8939 ETVIEKPTDALQITKEEVSSEAKTELSIPKAVREDKGTIYVASNRLGSVRNVAVEY 8998
 QY 6 ----- 5
 Db 8999 DRSPPRNLAVTDIKAESCYLTWDAPLNGSEITHYIDKRDASRKAEMEEVNTAVE 9058
 QY 6 ----- 5
 Db 9059 KRYGIMKLIPNGOEFRRVAVNKYGISDECKSDKVIYIODEPYRLPGPPKPYLARTKGM 9118
 QY 6 ----- 5
 Db 9119 LVSMTPLDNGSPITGYWLEKREBSPYWSRVSRAPITKVGKGVFNVRPLBVGKYO 9178
 QY 6 ----- 5
 Db 9179 FRAMAINAGIGPSESDPEVAGDPIFPGPSPCEVADKTSISLGWKPAPKDGSP 9238
 QY 6 ----- 5
 Db 9239 IKGIYEMOEGTDMKRVNEPDKLITTCCEVVPNLKELRYFRVAVNAEGESEPSDT 9298

Qy 6 ----- 5
Db 9299 TGEIPARDIOEPEVFIDIGAOLVCKAGSOIRIPAVIKGRPTPKSWEFDGAKKAMK 9358
Qy 6 ----- 5
Db 9359 DGVHDIEDAOLETAENSSVILLIPECKRSHTGKYSITAKKAGOKTANCRCVKMDVGP 9418
Qy 6 ----- 5
Db 9419 KDLKVSIDITRGSCLSMKMPDDOGDRLKGVIEKRTIDKAMTKVNPDCGSTTFVVDL 9478
Qy 6 ----- 5
Db 9479 LSEQOYFVRVAENRFGIGPPVETIORTADPIYPPDPPIKILIGLITKNTVHLSMKP 9538
Qy 6 ----- 5
Db 9539 KNDGSPVTHIYECIAMDPGTGCKEAMROCNKRDVELOFTVEDLVEGGEYEFVRKAVN 9598
Qy 6 ----- 5
Db 9599 AAGVKSATVPCDCORPMPISIDLKEFMEVEEGTNVIVAKIKGVPEPTLTWFKAP 9658
Qy 6 ----- 5
Db 9659 KKPDKNEPVLYDTHVNKLAVDDCTTIVIPQSRSDTGLYITAVNNLGTSKEMRLNLV 9718
Qy 6 ----- 5
Db 9719 RPPGPVPIKIFESYADQMTLSWPPKDDGSKITNVIEKREANKRTVHVSERPECT 9778
Qy 6 ----- 5
Db 9779 YTIPLKEGHEVYRIMAKQNKYIGIEPLDSEPTARNLFSVGPADPKPVSVTRNSMTV 9838
Qy 6 ----- 5
Db 9839 NMEEPEYDGGSPVTGYWLEMKDITTSKRWKRVNRPDKAMTLGVSRYKVTGLIESDYOF 9898
Qy 6 ----- 5
Db 9899 YAINAGVGPASLPSPDAPARPIAPGPPEPKVTMTKSSADLEMSPLKDGSKVTGY 9958
Qy 6 ----- 5
Db 9959 IVEYKEGKEEMEKDKVEYRGTGLVVTGLKEGAFYKFRVSAVNIAGIGPGEVTVIEM 10018
Qy 6 ----- 5
Db 10019 KDRLVSPDLQLDASVRDRIRYVHAGVIRITAIYVSGKRPPTVTVMNMNERILLPEATLET 10078
Qy 6 ----- 5
Db 10079 ISSSMVINKQQRSHQGVYSLAKNAGEKRTIIVDVLDPGVGPPLAHNLITNSCKL 10138
Qy 6 ----- 5
Db 10139 TWFSPEDDGSPITNVIEKRESDRAMPPTVTYTRQNAVQGLIQKRAVFRRIAENS 10198
Qy 6 ----- 5
Db 10199 IGMGPVETSEALVIREPITVPERPEDLVEKVTAKNTVTITMPPKXDGSEIINVLAS 10258
Qy 6 ----- 5
Db 10259 RLIGTEKHFVTNDNLISRKTYFKGLKEGDTYEVRSANVIGQKPSCTKPICTCKDEL 10318
Qy 6 ----- 5
Db 10319 APPTLHIDPDKLTIKVGAFALIGRYSKPKRVSMFKDEADVLDDRTHTIKTPATLIA 10378
Qy 6 ----- 5

Db 10379 LEKIKARSDSGKCYVVENSTGSRKFGQVNVVNDHPGPVPVSEDEVTKDYMIYSKMP 10438
Qy 6 ----- 5
Db 10439 PLDDGSKITNTYIEKKEVKDVMPEVTSASAKTCKYSKLEGRDIYFRIHAENLYGIS 10498
Qy 6 ----- 5
Db 10499 DPLVDSMKAKDRFRVPDAPDQPIYETVTKDSALVTWKNPBDGKPIITNYIIEKREYMSK 10558
Qy 6 ----- 5
Db 10559 RMARVTKDPIHPYTKFRVPDLLEGQYEFVSAENEGIGDPSPPSKPYFAKDIAPKSP 10618
Qy 6 ----- 5
Db 10619 PYNPEAIDTTCNSVDLTWOPPRHDGSKILGYIVEYQKVGEEMRRANHTPESCPETRYK 10678
Qy 6 ----- 5
Db 10679 VTGLRDGQTYKFRVLAVNAGESDPAHVPPEVLYKDRLEPPELLIDANMARQHIKVDGT 10738
Qy 6 ----- 5
Db 10739 LRLSAILKGVPPPKYTWKKEDBAPTKARIDVTPYGSKLEIRNNAHEDGGLYSLVENPA 10798
Qy 6 ----- 5
Db 10799 GSKTVSVKVLVDKRGPPRDLVESEIRDSCTLTWKEPLDDGSVITNYVVERDVASAQ 10858
Qy 6 ----- 5
Db 10859 WSPLSATSKKSHPAKHLNEGQYLFVRAENQYRGPPVETPKRIKALDPLHPPGPKD 10918
Qy 6 ----- 5
Db 10919 LHHVDYKTEVSLVWKNKPDROGSPITGYLVEYQEGTQDWIKFKTYNLECVVTLQOG 10978
Qy 6 ----- 5
Db 10979 KYRFRVKAENIVGLPPTTIPICQEKLVPSVELDKLIEGLVKAQTTRPAILIR 11038
Qy 6 ----- 5
Db 11039 GVPVPLAKWTJDSEIKTDEHYTETDNFSSVLTIKNCLRDTGERTYQTVSMAAGSKTYA 11098
Qy 6 ----- 5
Db 11099 VHLTVLDVGPPTGINILDTVPEHMTISWOPPKDDGSPVINTYVERKQTRKDTWGVYS 11158
Qy 6 ----- 5
Db 11159 SSSSKTKLIPHLOKGEVYFRVRAENKIGVPLDSTFTYAKHKFSPPSPGKRVVTDI 11218
Qy 6 ----- 5
Db 11219 TENAATVSWTLPKSDGSPITGYMERREVTGKWRVANKTPLADLKFRVTGLYGSNTYEF 11278
Qy 6 ----- 5
Db 11279 RYFAENLAGLSKPSSPDIKACRPDKPPPIKPKDOKSRETADLVWTRPLSDGSP 11338
Qy 6 ----- 5
Db 11339 LGYVEQCKPGTAQMNRIKNDKELIRQCAFVRYPGLIEGNEYFRIRKANIYVGEPRELAE 11398
Qy 6 ----- 5
Db 11399 SYLANDIHPPEVELDVCBVDITVRVQGTIRILARVGRBPDITWTKESGVLVREKRV 11458
Qy 6 ----- 5

Db 11459 DLIDDLRVELQIKEAVRADHGKIIISAKNSSGHAOGSALVNLDRPGCONLKVTVTK 11518
QY 6 ----- 5
Db 11519 ENCTISWENPLDNGSEITNEIYERKPNOKGWSIVASDVTKRLIKANLANNEYYFRVC 11578
QY 6 ----- 5
Db 11579 AENKVGCPITETKTPILAINPDRPGEPENLHADKGKTFVYIKMRPRDYDGGSPMLSY 11638
QY 6 ----- 5
Db 11639 HVERRLKSGDDMERVHKSGIKETHYVDRCVENOIYEFVQTKNEGSDMWKTEEVYVK 11698
QY 6 ----- 5
Db 11699 EDLOKPVLDLKLGVLTAKADTIRLEAGVGRKPPREAVMTKDKADFTLRSPRKIDTR 11758
QY 6 ----- 5
Db 11759 ADSSKFSLTAKRSDGKYYVLTATNTAGSFVATVNLKPGFVRLKIVDSSDCTV 11818
QY 6 ----- 5
Db 11819 CMDEPEDGGEIQNYLLEKCEKRWMTYSATVLTPTVTVRLLEGNEYIFRVAENK 11878
QY 6 ----- 5
Db 11879 IGTGPTTESKRVIAKTYDKRGRDPREPVTKVSKEMTVVWNPPEYDGSITGYLEKK 11938
QY 6 ----- 5
Db 11939 EKHSTRWVNVKSAIPERRMKVQMLPDHEXQFRVKAENEIGESPISRVAKADPIE 11998
QY 6 ----- 5
Db 11999 PPGPTNRVVDTKHSITLGMGRVYDGGAPITIGYVEMRPKITADASPDEGWRKCNAAA 12058
QY 6 ----- 5
Db 12059 QLVKREFVTSLDENOEFEFRVCAONOVGIGRAPALKEAIKPKILPPEIDLDASMRKL 12118
QY 6 ----- 5
Db 12119 VIVRACPIRLFAIVGRPAKVTWRKVYIDNVVRKGQVLDVTMAFLVIPNSTRDSDGK 12178
QY 6 ----- 5
Db 12179 YSLIIVNAGEKAVFVNVRLDTPGPVSDLVSDVTIKTSCHVSNAPPENDGGSQVTHIV 12238
QY 6 ----- 5
Db 12239 EKREADRKTWSGVNPEVKTSFHVTLNVPNGEYFRVTAVNEGPGVTDVPRVLAADP 12298
QY 6 ----- 5
Db 12299 LSEDPKRLATEMTKNSATLAMLPLRLDGCANIDGYIISYREEQPADRWTEYSVVKD 12358
QY 6 ----- 5
Db 12359 LSLVVTGLKEGKKYKFRVAANNAVSVSLPREAEGVYEAKEDLLPKILMPQIITIKAGKK 12418
QY 6 ----- 5
Db 12419 LRIEHAHYGKPHPCMKMKGEDEVVTSHLAVHRADSSSILIRDVTRKDSGYSLTAEN 12478
QY 6 ----- 5
Db 12479 SSGTDTQIKIKVYVNDAPGPPOPPEDISDIDADACSLWHIPLLEDGGSNTINYYEKCDVS 12538
QY 6 ----- 5
Db 12539 RGDWVTALASVTKTSCRGKLLPGQETIFRVAENRFGISEPLTSPKMNVOFPFGVPSER 12598

QY 6 ----- 5
Db 12599 KNAKRTKVNKDCIFVAMDRPDSGSPITIGYLIEKKENSLIMVANDTLVRSITEPCAG 12658
QY 6 ----- 5
Db 12659 IVEGLEYSFRYALUNKAGSSPPSKPTEVYTARMVPDPGKPEVIDVTKSTVSLIWARPKH 12718
QY 6 ----- 5
Db 12719 DGGSKLIIGYVEACKLPGDKWRCNTAPHOIPEEYTAGLEKAQYOFRAIARTAVNIS 12778
QY 6 ----- 5
Db 12779 PPSPSDPVTIILAENVPRPIDSVAMKSLITVYKAGTNCIDATVGRKMPYVSKKDCGL 12838
QY 6 ----- 5
Db 12839 LKPAEGIKMAMONLCTLELFSVNRKSGDYTITAEKSSGSKATIKLVLDPGPASV 12898
QY 6 ----- 5
Db 12899 KINKMTSDRAMLSWEPPLDGGSEITNTIVDKRETSRPNMAQVSATVPITSCVYEKLEIG 12958
QY 6 ----- 5
Db 12959 HEYQFRICAKNKYGVDPVFTEPATAKNPYDPPGRCDPPVISNITKDHMTVSWKPPADG 13018
QY 6 ----- 5
Db 13019 GSEITGYLLEKRETOAVNMTKVNRKPIERTLKATGLOGETEYERVAIINKAGPKPSD 13078
QY 6 ----- 5
Db 13079 ASKAATARDPQYPPAPPAKPYDTITRSSVSLSMGKRAYDGGSPITIGIYEVKRADSDNW 13138
QY 6 ----- 5
Db 13139 VRCNLPONLOKTRFEVYTGLEDTOYQFRVYAVNKGISDPDVPDKHYPKDILIPPEGH 13198
QY 6 ----- 5
Db 13199 DADLRFTLILRAGVTNRLIYPVKGPRPKITWSKPNVNLDRIGLDISTDFEFLRCEN 13258
QY 6 ----- 5
Db 13259 VNKYDAGKYIITLENSCGKKEYTIYKVLDTPGPPIVNTVKEISKDSAVTWMEPPLIDGG 13318
QY 6 ----- 5
Db 13319 SPITINTVOKRDAERKSMSTVTECSKTSFRVPNLEBGSYFFRVFAENEXGIDPGETR 13378
QY 6 ----- 5
Db 13379 DAVKASQTPGPVVDLKVRSVSKSSCSIGMKKHSDDGSIIGIYVVDLFEENKMRVWKS 13438
QY 6 ----- 5
Db 13439 LSLQYSAKDLTEGKEYTFRVSAENENGEGTPEBITVVARDDVAVAPDLKGLDCLYLAK 13498
QY 6 ----- 5
Db 13499 ENSNFRKIPIDKGRAPASVSWKKGEDPLATDRVSVSSAVNTTLLIYDQKSDAGKYTI 13558
QY 6 ----- 5
Db 13559 TLKNVAGTEKGTISIKVVGKPGIPTGPIDDEVTAEMTLKNAAPKDDGGSITNTYLEK 13618
QY 6 ----- 5
Db 13619 RDSVNNKWTWCASAVOQKTFTRVTRLHEGMEYTFRVSANENKYGVBGLKSEPIVARHPDV 13678

QY 6 ----- 5
Db 13679 PDAPPPNIYDVHRDVSLSLWTDPKTGGSPITGYHLEFFERNSLMKRANKPIRMRDF 13738
QY 6 ----- 5
Db 13739 KYVGLTGBLEEFYVMAINLAGVKSPLPSEPVVALDPIDPPGKPEVINITRNSVTLIWT 13798
QY 6 ----- 5
Db 13799 EPKYDGHKILGYIYERKDLPSKSMKANHVNECAFTVTLVEGKYEFRIRAKNTAG 13858
QY 6 ----- 5
Db 13859 AISAPSESTETIICKDEYEAPTVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSK 13918
QY 6 ----- 5
Db 13919 AGKDIRSDITQITSTPSSMLTIKATRKDAGEYITATNPGETVENVKTVLVDYPCP 13978
QY 6 ----- 5
Db 13979 PGVEISNVSAREKATLTWTEPLEDGGSPIKSYILLEKRETSRLMTVYSEDIGSCRHVATK 14038
QY 6 ----- 5
Db 14039 LIQNEIYIFRVSANHYKGEPVQSEPVKAVDRFGPPPEKPEVSNVTKNTATVSMKRP 14098
QY 6 ----- 5
Db 14099 VDDGSEITGYHYERREKSLRWVRAIKTPVSDLRCKVYGLQGSYIEFRVSAENRAGIG 14158
QY 6 ----- 5
Db 14159 PPEASDSVLAKDAAYPPPPSNPHVDTTKKASLAWKPHYDGLLEITGYVEHQVYG 14218
QY 6 ----- 5
Db 14219 DEAMINDTGTALRITQFVVPDLQTKKYNFRISAINDAGVGPVAVIPEVEIVERMADP 14278
QY 6 ----- 5
Db 14279 FELDAELRRLVVRAGISIRIFVPIKGRPAEYTWTKONLNKRNANIENTESFTLLIIP 14338
QY 6 ----- 5
Db 14339 EGNRYDTGKFVMTIENPAGKSGFVNVRLDTPARSPQLRPTDITKDSYTLHMDLPLIDG 14398
QY 6 ----- 5
Db 14399 GSRITNVIYEKREATRKSSYSTATTKCHKCYKVTGLSEGCETFRVMAENEGIGEPJET 14458
QY 6 ----- 5
Db 14459 TEHVKASEAPSPDSINIMDTKSTVSLAMPKPKHDGSKITGYVLEAQRKSDQWTHIT 14518
QY 6 ----- 5
Db 14519 TYKGLECVVRNLTGESEYTFQVMAVNSAGRSAPRESRPVIREQTMPELIDLRGITQKLY 14578
QY 6 ----- 5
Db 14579 IAKAGDNIVKELPVLRPRPYTWKKGDOILOKOTQVNFETATSTIILINECVASDSGP 14638
QY 6 ----- 5
Db 14639 YPLTANIVGEGDVITIQVHDIKPPPTGPIKFEDEVSDVFMSDPENDGVPISNVY 14698
QY 6 ----- 5
Db 14699 VEMKQDSTTWELATTVIRITTYKATRLTGLGYOFVKAQNRVGVGSGITSAMIVANP 14758
QY 6 ----- 5

Db 14759 FKVPBPPTQYVAVTKDSMTISMHEPLSDGSPILGYHYERKENGILMQTVSKALVPG 14818
QY 6 ----- 5
Db 14819 NIFKSSGLTDGIAYEERYIAENMAGSKSPSEPMALDPIIDPPGKVPPLNITRHVTL 14878
QY 6 ----- 5
Db 14879 KNAKPEYTGFKITSYIYERKDLPNGRMLKANFSNILENFTVSGLTEDAAVEFRVIAKN 14938
QY 6 ----- 5
Db 14939 AAGAISPPSPSDATTCRDDVEAPKIKVDVKFDTVILKAGEAPRLADVSGRPPPTMEW 14998
QY 6 ----- 5
Db 14999 SKDGKLEBGTAKLEIKIADSFNLVKNKSTRDSGAYTLTATNPGFAKHIFNVKVLDRP 15058
QY 6 ----- 5
Db 15059 GPPEGPLANTVEYTSKCVLSWFPPLDDGAKIDHYVQKRETSRLAMTNVASEVQVTKLK 15118
QY 6 ----- 5
Db 15119 VTKLKNEYIFRVAVNVKYGVEPLESEPVLAVPYGPDPKPEVTTITKDSMVCW 15178
QY 6 ----- 5
Db 15179 GHPDSDGSEIINIVYERDKAGQRMKCNKTLTDLRYKVSGLTEGHEYFRMAENAA 15238
QY 6 ----- 5
Db 15239 GISAPSPSPFYKACDITFKBPENPRVLDTSRSSISIANKPIYDGSSEITGYMVEIA 15298
QY 6 ----- 5
Db 15299 LPEEDEMOITVPPAGIKATSYITIGLTENOBYKIRIYAMNSEGLGEPALVPTPKAEDRM 15358
QY 6 ----- 5
Db 15359 LPELELDADLRKVYTIRACCTLRIFVPIKGRPDEVKWADHGBSLDKASIESASVTL 15418
QY 6 ----- 5
Db 15419 LTVGNRBFDSGYILLIENSNGSAFVNVRLDTPGPQDLKVEYTKTSVTLTMDPP 15478
QY 6 ----- 5
Db 15479 LLDGSKIKNVIYERKSTRAYSTVATNCHKTSWKVDLOEGCSYFRVLAENEYIGL 15538
QY 6 ----- 5
Db 15539 PAETASVKASBRPLPCKITLMDVTRNSVLSWEKPBHDGSHLGYIYEMQTKGSDKW 15598
QY 6 ----- 5
Db 15599 ATCATVKTETATIGLIOGEYSFRVSAQNEKGISDPRQLSVPYIANDVILPFAKILFN 15658
QY 6 ----- 5
Db 15659 TPTVLAGDLKVDVPEIGRPTPAVTWKNVPLKOTTRVNAESTENSLTIKDACHEDV 15718
QY 6 ----- 5
Db 15719 GHYVVKLINSAGEALIELTLNIVLIDKPGPTGPKMDEVTAISTLSMGPKRYDGSINN 15778
QY 6 ----- 5
Db 15779 YIYERKDTSTTWOIVSATVARTIKACRLKTCGEYOFRIAAENRYGKSTYLNSEPTVAO 15838
QY 6 ----- 5

```

Db 15839 YPEKVPGPPTPVYLLSRDSMEVQWNEPISDGSRVIGYHLERKERNILMWKLNKTPI 15898
QY 6 ----- 5
Db 15899 PQTKEKTTGLEEGVEEERVAENIVIGKPSVSECYARDCPPGPEALITRNSV 15958
QY 6 ----- 5
Db 15959 TLOMKKPTDGSKIINGYIYEKKELPEGRMMKASFTNIIDTPEVYGLVEDHREFRVLA 16018
QY 6 ----- 5
Db 16019 RNAGVSESESTGATTADEVDPPRISMDPKYDTIVVHAGESEKVDADYKDPIT 16078
QY 6 ----- 5
Db 16079 QWIGQDELSTARLEIKSIDFATSLSVKDAVRDSCNTLLKKNVAGESVTVNKKVLD 16138
QY 6 ----- 5
Db 16139 RPPPEGPVVISGVTAEKCTLANRPLQDGSIIINIVERRERSRLVWTVVANVQTL 16198
QY 6 ----- 5
Db 16199 CKVTKLEGNEYFRIMAVNKYGVGEPLSEBPVYAKNPVVPDPAKAEVTVTKDSMT 16258
QY 6 ----- 5
Db 16259 VMERPADGSEILGYLEKRDKEGIMWTRCHRLIGELRLRTGLLENIDYERVAEN 16318
QY 6 ----- 5
Db 16319 AAGSESPSPAYQACDPIYKPGPPNNPKVIDITRSSVFLWSKPIYDGCCEIOGYI 16378
QY 6 ----- 5
Db 16379 KCDVNVGEMTCTPPTGINKTINIEVEKLEKHEYNFRICAINAGVEHADVGPPIVEE 16438
QY 6 ----- 5
Db 16439 KLEAPDIDLELRKIINIRAGSLRFLVPIKGRPEVKWKVDGEIRDAIIDVTSF 16498
QY 6 ----- 5
Db 16499 TSLVDNVRYSKTYLTLENSSGTSAPVTVRVDTPSPVNLKYTEITKDSVSITWE 16558
QY 6 ----- 5
Db 16559 PPLDGGSKIKNYIYERREARRKYAAVNTNCHNSMKIDQLDGCSTYFRVTAENEGY 16618
QY 6 ----- 5
Db 16619 GLPAOTADPIKVAEVPQPGKITVDVTRNSVLSWTKPEHDGSKIIQYIVEMQAHSE 16678
QY 6 ----- 5
Db 16679 KMSCARVKSLOAVITNLTOGEETLFRVAVANEKGRSDPSLAVPIYAKDLVIEPDVKA 16738
QY 6 ----- 5
Db 16739 FSSSVQVQDGLKMEVPISGRPPLITWTKDGPLKQOTTRINVTDSLDTLSIKETHKD 16798
QY 6 ----- 5
Db 16799 DGGGYITVANVGOQTASIEITLDRKDPKGVKFDVSAESITLSWNPPLYTGCCQI 16858
QY 6 ----- 5
Db 16859 TNYIVQKRDITTTVMWVVSATVAARTLLKVTKLGTGTEQRIFAENRYGOSFALESPIV 16918
QY 6 ----- 5
Db 16919 AQYVPEKPGPPTPATAISKDSMWIQWHEPVNNGSVIGYHLERKERNILMTKVNT 16978

```

```

QY 6 ----- 5
Db 16979 IITHDQFKAQNLBEGIEIEFRVTAENIVGVKASKNSCXYARDPCDPGPEPIWKRN 17038
QY 6 ----- 5
Db 17039 EITLQWTKPVYDGSIMINGYIYEKRDLPDGRMMKASFTNIETQFTVSGLTEDQRYEFRV 17098
QY 6 ----- 5
Db 17099 IAKNAGAIKSPDSTGPTTAKDEVELPRISMDPKFDTIVVNAGETFRLEADVHGKPL 17158
QY 6 ----- 5
Db 17159 TIEWLRGDKIEBSARCEIKNTDFKALLIVKDIRIDGQYILRASNVAGSKSEPVNKKV 17218
QY 6 ----- 5
Db 17219 LDRPPEGPVQVGTSEKCSLTWSPPLQDGSISHYVEKRETSRLANTVVAASEVYT 17278
QY 6 ----- 5
Db 17279 NSLKYTKLEGNEYFRIMAVNKYGVGEPLSAPVLMKNPVLQPPPKSLEVTNIANDSM 17338
QY 6 ----- 5
Db 17339 TVCWNRPDGSGSEILGYIYERKDRSGIWMKCNRRITDLRLRVYGLTEDEHEFEFRVA 17398
QY 6 ----- 5
Db 17399 ENAAGVESPATVYIYKACDPVFKPDPPTNAHIUDTNTNSITLWAKPIYDGCSEILGY 17458
QY 6 ----- 5
Db 17459 VELCKADEEMQIVTPQGLARTFREISKLEBHQYKIRVQALNKVGLGEATSVPGTKP 17518
QY 6 ----- 5
Db 17519 EDKLEPDLDELSELRKGIIVRAGSARLHIFPKRPMPEITWSREBETDKVQIEKV 17578
QY 6 ----- 5
Db 17579 NYTOLSIDNCDNDACKYLLKLENSGSKSAPVTVKVDTPGPONLAVKEVRKDSAPLV 17638
QY 6 ----- 5
Db 17639 WEPLIIDGAKVKNYIDKRESTRKAYANVSSKSTSEKVENLTGCAIYFRVAENEF 17698
QY 6 ----- 5
Db 17699 GVGVPVEIVDAVKALEPPSPGKITLTDVSTQTSASLAMEKPEHDGSKVLYGVEMQPKG 17758
QY 6 ----- 5
Db 17759 TEKMSIYAESKVCNANVTGLSSGQYQFRVKAYNKSKSDPVLGVPIAKDLTIQPSLK 17818
QY 6 ----- 5
Db 17819 LPENTYSIOAGEDLKEIPVIGRPRNISWVDGEPLKQOTRVNVEATSTVLHIKSGN 17878
QY 6 ----- 5
Db 17879 KDDFGKITYTAINSAGTATENSVIYLEKPPGVKREDEVASADPVVISWEPPAYTGCC 17938
QY 6 ----- 5
Db 17939 QISNVIYERKRDITTTVMWVVSATVAARTLLKVTGTEYQRIFAENRYGOSAPLDSKA 17998
QY 6 ----- 5
Db 17999 VIVQYPEKPGPPTPATAISKDSMWIQWHEPVNDGTTKIIYHLERKERNILMTKVN 18058

```

QY 6 ----- 5
Db 18059 KTIPIODTKFTGTDEGLEEFKSAENIVIGIKRPSKVSCEGFVARDPCDPGPEAIYIT 18118
QY 6 ----- 5
Db 18119 RNNVTLLAKKKAPAYDGSKITGYIEKKDLPGRRMKASFTNVLETFETVSGLEDORXYE 18178
QY 6 ----- 5
Db 18179 RVIAARNAGNFSEPSDSSGALTARDEIDAPNASIDPKYKDIYVHAGETVLEADIRGKP 18238
QY 6 ----- 5
Db 18239 IPDVVMSKDGLEBETARMEIKSTIOKTLVVKDCIRTDGQYILKLSNVGTSKIPIT 18298
QY 6 ----- 5
Db 18299 VKVLDREPSPEGLKYGTVAEKCYLAMNPPLDGGANISHYITTEKRETSRLSTQVSTE 18358
QY 6 ----- 5
Db 18359 VOALNTYKTKLLPGNEVIFRVMAVKNKYGIGEPLESGPYTACNPYKPPGPSTPEVSAITK 18418
QY 6 ----- HSMEXR 11
Db 18419 DSMVVTMARPVDDGTEIEGYILEKRDKEGVRMTCKNKKTLJDLRLRVGTGLTEGHSYEF 18478
QY 12 ----- 11
Db 18479 VAAENAGVGEPSPEPVYRACDALYPPGPSNPKYDTSRSSVSLAMSKPIYDGGAPVK 18538
QY 12 ----- 11
Db 18539 GYVVEVKEAADDEWTTCTPTLGOKQFVYKLEKTEYENRICALINSEGVGEPAITLPGS 18598
QY 12 ----- 11
Db 18599 VVAOERIEPPELIDADLRKVVVILASATYLRLEVYIKGRPEPEVKEKAGILTDRAQIE 18658
QY 12 ----- 11
Db 18659 VTSFTMLVIDNVRFDPSGRNLTLENNSGSKTAFVAVRILDSAPVNLIREVKDSY 18718
QY 12 ----- 11
Db 18719 TLSNEPPLIDGAKITNYIEKRETRKAYATITNNCTKTFFRIENLOEGCSYFFVLAS 18778
QY 12 ----- 11
Db 18779 NEYIGLPAETTEPVKSEPLPPGRVTLVDVTRNATITWEKREPESDGSKITGYVVEMO 18838
QY 12 ----- 11
Db 18839 TKGSEKSTCTQVKTLEPATISGLAGEEYFVAAVNEKGRSPROLGVPYIARDEIFIK 18898
QY 12 ----- 11
Db 18899 SVELPHTFENVKAREQKIDVPRGRPOATVNNRKDGQTLKETTRVNVSSKTVTSLSIK 18958
QY 12 ----- 11
Db 18959 EASKEDGYELCVNSAGSIYPIITIVLDRGPRGPIRIDEVSCDSITISNPPETYG 19018
QY 12 ----- 11
Db 19019 GCOISNIVYEKETTSTTHIVISOAVARTSIKIVRLTGTSEYQFVCAENRYGKSSYSE 19078
QY 12 ----- 11
Db 19079 SAYVAEIPSPGPGPIPKVYVHATKSTMLVTWOPVNDGGSRVIGYHLEVKERSILWSE 19138
QY 12 ----- 11

Db 19139 ANKLLADQVVKVSGLEGLMEYERYAENIAGICSKSCPEVPARDPCDPGPEVTN 19198
QY 12 ----- 11
Db 19199 ITRKSVLKMSPHYDGAKITGYIERRELPDGRMLKCNNTYNIQETFEVTELEDQRY 19258
QY 12 ----- 11
Db 19259 EFRVAPARNADSVSESESTGPIIVKDYEPPEVMDVFRDVIYVKGAVLKINADIAG 19318
QY 12 ----- 11
Db 19319 RPLPVISMADGIEIERARTEIISTDNHTLLVKDCIRBDTGGQVYVLTKNVAGTRSAV 19378
QY 12 ----- 11
Db 19379 NCKVLDPGPPAGPLEINGLTAEKCSLSWGRPOEDGADIDYHRRKRETSHLAWTICBG 19438
QY 12 ----- 11
Db 19439 ELQMTCKYTKLLKGNBYIFRVYGVNKGVEPLESVAIKALDPTVPSPTSLTTSVT 19498
QY 12 ----- 11
Db 19499 KESMTLCWSPESDGSSEISGYIIEERKNSLRWVAVNNKPYDLRYKSTGLREGCEYEX 19558
QY 12 ----- 11
Db 19559 RYVAENAGLSLPSSTSPILRAEDPVFLPSPSKPIVDSKTTITIAMVKKPLFDGAPI 19618
QY 12 ----- 11
Db 19619 TGYTVEYKKSDDDMKTSIQSLRGTEYITISGLTGAETVFRVKSVMKVGASDPSDSDPQ 19678
QY 12 ----- 11
Db 19679 IAKEREEDPLFDISEMRKTLIVKAGASEMTVPFGRGVPVNLMSKPDJDLRTAYYDT 19738
QY 12 ----- 11
Db 19739 TDSRTSLTEMANRNDSGYITLLIONVLSAASLTIVKVLDPGPPNTIVODVTKESAV 19798
QY 12 ----- 11
Db 19799 LSWDVENDGAPVKNYHIEKREASKKAWSVTNNCNRLSYKVTMLQEGAIYFVSGEN 19858
QY 12 ----- 11
Db 19859 EFGVGIPAEKTEGVKITEKPSPEKLGVTISKDSVSLTWLKEPHDGSRIYHYVVALE 19918
QY 12 ----- 11
Db 19919 KGOKMVKCAVAKSTHHVVSGLRENSEYFFRVFAENQAGLSDPRELLPLVIKEQLEPPE 19978
QY 12 ----- 11
Db 19979 IDMKNPSTHYVVRAGSNLKVDPISGRPLPKVTLSRDGVPLKATMRNTEIETANLITN 20038
QY 12 ----- 11
Db 20039 LKESVTADAGREYITANSSGTTKAFINIVLDRBPPTGPVVISDITEESVTLKWBPPK 20098
QY 12 ----- 11
Db 20099 YDGSQVNTYIILKKRETSNAVTEVSATVARTMAKVMKLTIGEEYQFRIKAKENREGISDH 20158
QY 12 ----- 11
Db 20159 IDSACVTVLKYPTTPGPPSTPWTNVTRESITVGVHEPVSNGSAVGVYHLEMDRNSIL 20218
QY 12 ----- 11

Db 20219 WOKANKLIVITTHFKVTTISAGLIYEEFVYAENAGVSKPSHPSEPVLAIDACEPPHNR 20278
 QY 12 -----TPD----- 11
 Db 20279 ITDISKNSVLSMOOPAFDGSKITTYIVERDLPGDRMTKASTNTNETOPTISGLTON 20338
 QY 12 ----- 11
 Db 20339 SQEPRVAFARNAGVSSINPSEVGPITCIDSYGYPVLDLPLETEYVYKAGTSVKLRAG 20398
 QY 12 ----- 11
 Db 20399 ISGRPAPIEYKDDKELQNALVCVENTYDLASILIKADRLNSGCEYELKRNAMASAS 20458
 QY 12 ----- 11
 Db 20459 ATIRVQIILDKRPGPGRPLEFTVTAKEITLLMRPPADGGAKITHYIYERKETSRYVWSM 20518
 QY 12 ----- 11
 Db 20519 VSEHLEECIITTTKIKNGEYIFRVAVNKYIGIEPLESDSVAKNAFVTPGPGIPEVT 20578
 QY 12 ----- 11
 Db 20579 KITNSMTVWSRPIADGSDISGYFILEKRDKSLGMEFKYKJETIRDTROKVTGLTENS 20638
 QY 12 ----- 11
 Db 20639 YQYRCVAVNAAGCPSPSPSEFKAADPIDPPCPAKIRIADSTKSSITLQMSKPYVDG 20698
 QY 12 ----- 11
 Db 20699 SAVTGYVEIROGEEEMTIVSTKGEVFTTEYVSNLPGVNYFRVSAVACAGCEPIE 20758
 QY 12 ----- 11
 Db 20759 MNEPVQADILEAPEIDDLALRTSVIAKAGEDYQVLIPEKGRPPYVYWRKDEKNLGS 20818
 QY 12 ----- 11
 Db 20819 ARYSIENTDSSLLIIPQVTHNDGKYLITLENGVEPKSTSVKVLDPFAOQKLOVK 20878
 QY 12 ----- 11
 Db 20879 HVSNGVTLLMDPPLIDGSPINVIIEKRDATKRTWSVSHKCSNPSFKLIDSEKTPF 20938
 QY 12 ----- 11
 Db 20939 FFRVLAENEIGIEPCETTEPEYKAEVAPIRDLMSKDDSTSVILSWTKPFDGGSYIT 20998
 QY 12 ----- 11
 Db 20999 EYVVERKKGEGTWSHAGISKTCEIVSOLKEQSVLEFRVAPKANKGLSDPVTIGPITVK 21058
 QY 12 ----- 11
 Db 21059 BLITPEVLDSDIPCAQVTVRIGNVHLELPYKGPKPSISMLDKGLPLKESEFVRSKT 21118
 QY 12 ----- 11
 Db 21119 ENKITLSIKNAKKEGKTYIILDNACRIAPITVITLGPSPKPGIRDEIKADSVI 21178
 QY 12 ----- 11
 Db 21179 LMDVPEDNGGGEITCYSTIEKRETSQTNMKMVCSSVARTTFKPNLVDAEYQFVRBAEN 21238
 QY 12 ----- 11
 Db 21239 RYGVSOPLVSSIYAKHOFRIIDGPGKPVYIYVSDGSLTWDAVYDGGSEVTFHVEK 21298
 QY 12 ----- 11
 Db 21299 KERNSILMOKVNTSPISGRETRATGLVEGLDYQFVRYVAENAGLSSPSDPKFTILAVSPV 21358

QY 12 -----TPD----- 14
 Db 21359 DPGGPFIDYIDVYREHITTLKMPNPLDGGSKIVGSIIEKROGNERVNRONFTDVSQOYTV 21418
 QY 15 ----- 14
 Db 21419 TGLSPDRYEFRILARNAVGTISPPSQSGIIMTRDENVPIVERGPEYFDGLIKSGES 21478
 QY 15 ----- 14
 Db 21479 LRKALVQGRPVRYTWFKDQVEIEKRMMETITNLGSLTFVADATDHRGVTVEAKN 21538
 QY 15 ----- 14
 Db 21539 ASGSAKAEIKVQDTPGKVGPPIRFTNITGEMKTLMDAPLNDGCAPITHYIIEKRETS 21598
 QY 15 ----- 14
 Db 21599 RLAMALIEDKCEAOSTAIAKLINGNEYOFVSAVKNKFGVREPLSDPVVAQIOYTVDPAP 21658
 QY 15 ----- 14
 Db 21659 GIPEPSNITGNSITLTWAPESDGSSEIQYILEREKSTRWVAVISKRPISETRKYT 21718
 QY 15 -----INP----- 17
 Db 21719 GLTEGNEVEFHVMAENAGVGPASGISRLIKCREPVNPPGPTVVKVTDTSKTVLSLMS 21778
 QY 18 ----- 17
 Db 21779 KPVFDGMEIIGYIEMCKTDLGDMHKVNAEACVTRTYTVDLQAGEEYKFRVASIANG 21838
 QY 18 ----- 17
 Db 21839 KGDSEVGTGIRAVDLRLAPELDIDANFKQTHVVRAGASIRLEIYVQGRPTVAVWSKD 21898
 QY 18 ----- 17
 Db 21899 SNLSLRADITHHTDSFSTLIVENCNRNDAGKYTLTVENNSSKSTFTVAVLDPGPGPI 21958
 QY 18 ----- 17
 Db 21959 TFRDVTGSAATLMDAPLIDGAGRIHHYVEKREASRSRMOYISEKCTROIKTVNDLAE 22018
 QY 18 ----- 17
 Db 22019 VPYFRVSAVNEVGSEPEYEMPEPIVATEQAPAPRRRLDVDTSKSAAVLAWLKPDDGS 22078
 QY 18 ----- 17
 Db 22079 RINGYLLEMRONGSDLMVAGHTKQLTFTVERLVEKTEYEFVAKAKNDAGSEPREAFSS 22138
 QY 18 ----- 17
 Db 22139 VTIKEPOIEPTADLTGITQOLITCKAGSPFTIDVPISGRPAKVTWKLEMRLEKEDRYS 22198
 QY 18 ----- 17
 Db 22199 ITTTKDRITLTVKDSMRGSGRFTLLENTAGKTFESVTVVIGRGPVTPGLEVSSVA 22258
 QY 18 ----- 17
 Db 22259 ESCVLWGEKDGGEITNYIYERKESGTAMOLVNSSVKRQIKVTHLTRYMEYSFRV 22318
 QY 18 ----- 17
 Db 22319 SSENRFVSKPLSAPITIAEHFPVPSAPTRREVIVHSANAMSIRKEEYHDGSKTIQY 22378
 QY 18 ----- 17
 Db 22379 WVEKKERNTILMWKENKVPOLCECNKYVTGLVBSLEVOFTYALNAGVSKASEASRPIMA 22438

QY 18 ----- 17
Db 22439 QNPVADAPRPEVTDVTRSTVSLIWSAPADGSKVGYIIBRKPVSEVGDGRMLKNCVTI 22498
QY 18 ----- 17
Db 22499 VSDNEFTYALSEGDTEFEFRVLAKNAGVISKSESTGPTVCRDEYAPPKAELDARLHGD 22558
QY 18 ----- 17
Db 22559 LVTRFAGSDVLDAVAGSKPEPKIITWKDKELDCEKVSQYTGKRTATVIRKCDSDS 22618
QY 18 ----- 17
Db 22619 GKTYLTYNKASGTRAVSYMVKVLDSPGCGKLTVSRYTQEKCTIAMSILPOEDGAEITHY 22678
QY 18 ----- 17
Db 22679 IVERRETSRLMWIVVEGCEPTLSYVTRLLKNNETIFRRAVANKYGGVPESEPIVARN 22738
QY 18 ----- 17
Db 22739 SFTIIPSPGIDPEVGTGKEHIIQWTKPESDGNESINYLDKREKESLMTIRVNDYV 22798
QY 18 ----- 17
Db 22799 YDTRLKATSLMEGCDYQPRYATVANAAGNSEPERSNEISGNEPSYTPGPPSAPRVDTTK 22858
QY 18 ----- 17
Db 22859 HSISLMTKPMYDGDITVGVLEMOEKDDQWYRVHTNATIRNTEFTVDLKKGQKYSF 22918
QY 18 ----- 17
Db 22919 RVAAVNWKMSSEVSEIAIEPEVPERIEIPDELADLKTVTIRAGASLRIMVSGRPP 22978
QY 18 ----- 17
Db 22979 PVITWKSOGIDLASRAIITTESISLIYDKVNRDAGKTTIEAENOSGRKSATVLKVY 23038
QY 18 ----- 17
Db 23039 DTPGPCSVKKEYSRDSYTIWEIPTIDGAPINNIYKREAMAFAFTVTTKSKTL 23098
QY 18 ----- 17
Db 23099 YRISGLVEGTMHFRVLPENITGIGECETSDAVLSEVPLVPAKLEVVDVTKSTVLIAM 23158
QY 18 ----- 17
Db 23159 EKPIYDGSRLTGVLEACKAGTERMMKVYTLKPTVLEHTVTSINBEOYLFRIRAOONEK 23218
QY 18 ----- 17
Db 23219 GVSEPRETVAVVODLRFVLTIDLSMPOKTIHVPAGRVELVPLAGRPAPASWFA 23278
QY 18 ----- 17
Db 23279 GSKIRESEKVTETHTKVAKLTIRETTIRDTGEYTLKKNVGTSTETIKVITILDKRGP 23338
QY 18 ----- 17
Db 23339 TGPRIKIDEIDATISITISWEPPELDGAPLSGYVVEQDRAHRPQWLVSSEVTRSTFKTR 23398
QY 18 ----- 17
Db 23399 LTEGNEVFRVAATNRFIGISYLOSEVIECRSSIRIPGPETILOIDVSRDQWLTWPP 23458
QY 18 ----- 17
Db 23459 EDDGSGVTCYIYERKEVRADRWVRVNVKPYVTMTFRSTGTLGLEYEHRVTAIINARSG 23518
QY 18 ----- 17

Db 23519 KPSRPSKPIVANDPIAPRPGKPNRPRTDITRTSVLSAMSVPDEGSKVTGILIEQKVD 23578
QY 18 ----- 17
Db 23579 QHEMTKONTPTKIREYITLHLPOGAERYRVLACNAGGPGPAEYGVTKYTEMLEYPD 23638
QY 18 ----- 17
Db 23639 YELDERTQEGIFVRQGGVIRLTIPIKGPPEPICWKTEGODISKRAMIATSETHLELVK 23698
QY 18 ----- 17
Db 23699 EADRGSGTYDVLNKKCKGKAYIKRVYIGSPNSPEGLEVDIOVRSVRSMRPADD 23758
QY 18 ----- 22
Db 23759 GGADILGYTLERREVRKAMYTIDSHVSGTSLVYKGLKENVEYHFRVSAENQFGISKPLK 23818
QY 23 ----- 22
Db 23819 SEEPVTPKPIPLNPPEPPSNPEYLDVTKSSVLSMSRPDGGSRVGYIERKETSTDK 23878
QY 23 ----- 22
Db 23879 VVRHNKTQITTMVTVTGLVPAEYQFRILQNDVGLSETSPASPEVYCKDFDKPSQPG 23938
QY 23 ----- 22
Db 23939 ELEIISKDSVTLQWKEPCDGGKEILGYWYEROSGDSAMKSKNERIKDKOFTIGL 23998
QY 23 ----- 22
Db 23999 LEATEYFRVPAENETGLSRPRTAMISITKLTSGAPGIREKMDVTTKIGEAOLSCQ 24058
QY 23 ----- 22
Db 24059 IYGRPLDIKMYRFGKELIOSRKYMSDGRHTLTVMTQEDBQVYTCIATNEVEVE 24118
QY 23 ----- 22
Db 24119 TSSKILLQATPOPHGYPKKEKTYGAVGSTLRHVMYIGRPVPMTWGOKLQNSENI 24178
QY 23 ----- 22
Db 24179 TIENTEHTLVKMNQKRTHAGKYKQVLSNVEGTVDAILDVEIQDKPFGPIVEAL 24238
QY 23 ----- 22
Db 24239 LKNSAVISWKPADDGGSWITNYVEKCEAKEGAEMLVSSAISVTCTRIVNLNENAGY 24298
QY 23 ----- 22
Db 24299 FRVSAONTGIDPLEVSSVYIKSPFEKPAKFTITAVTKDSCVAMKRPASDGAK 24358
QY 23 ----- 22
Db 24359 IRNYYLEKREKQNMWISVTEEIRETFESVKNLIEGLEEPRVCENLJGSESEMSISE 24418
QY 23 ----- 22
Db 24419 PITPKSDVPICQPHKEELRNINRYOSNAPLVCKVTGHPRIYKMYROGKEIADGLKY 24478
QY 23 ----- 22
Db 24479 RIOEFKGYHOLIASVDDATVYQVRATNGGSVSGTASLAEVPAKIHLPKTLBOMG 24538
QY 23 ----- 22
Db 24539 AVHALRGEVYSIKIPISGRKPDVITWOKGODLIDNNGHYQVITVRSFTSLVFPNGVERKD 24598
QY 23 ----- 25
-RGI-
!!!

Db 24599 AGFYVCAKNRFGIDOKTEVELDADVPDPGRGVKVSASRDSVNLTWTEPASDGSKITN 24658
 QY 26 ----- 25
 Db 24659 YIVEKATTAERMLRVGOARETRTYVNLNGKTSYQRYVLAENKFGLSKSPSEPTITK 24718
 QY 26 ----- 25
 Db 24719 EDKTRAMNYDEVDRETSMTKASHSSTKELYEKYMAIEDLGREGYVHRCVETSSKK 24778
 QY 26 ----- 25
 Db 24779 TYMAKFVKGTQOVLVKEKISILNIAHRNLIHLHESFESMEELVMIFFISGLDIFER 24838
 QY 26 ----- 25
 Db 24839 INTSAFELNEBEIVSYVQVCEALQFLSHNIGHFDIREPMIYQTRSSYIKIIEEQA 24898
 QY 26 ----- 25
 Db 24899 ROLKPGDNFRLLETFAPERYAPEVHOHDVSTATDKMSLGLTVYLLSGINFLAETNOQI 24958
 QY 26 ----- 25
 Db 24959 IENINAEYTFDEAEFAKEISIEAMDFVDRLVKERKSHMTASEALQHEWIKOKIERVSTK 25018
 QY 26 ----- 25
 Db 25019 VIRTJLKHRRYHTLJKRLNMYVSAARISCGAIRSQGVSAKVAKVASEIGVSGQIM 25078
 QY 26 ----- 25
 Db 25079 HAVGEEGHVKYCKIENYDOSTQVWYFGVROLENSEKEITYEDGVALLYVDITKLD 25138
 QY 26 ----- 25
 Db 25139 DGTVCCKVNDYGEDSSYAEFLVKGVRREVYDYCHRTMKKIKRRTDMRLLEPPEPTLP 25198
 QY 26 ----- 25
 Db 25199 LYNKTAAYGENVRGVITLVHPEPHVTWYKSGQIKPGDNCKYTFESDKGLYQUTINSY 25258
 QY 26 ----- 25
 Db 25259 TTDDAEYTVARNKYGEDSCAKLTVLHPPTDSTLRPMFKRLLANAECOGOSVCFE 25318
 QY 26 ----- 25
 Db 25319 IYVSGIPEPTLWKEDGQPLSIGPNIEIILHEGLDYALHINDTLPEDTGYRYVATNTAG 25378
 QY 26 ----- 25
 Db 25379 STSCQHLQVERLYRKQEFKSEHERHVQKQIDKTLMAELISGTESVPLTVAKREAL 25438
 QY 26 ----- 25
 Db 25439 REAAVLKPAVSTKTVKGFELEIEKKEERKLAMPYDVPPEPKYKQTTIEDQRIKQFV 25498
 QY 26 ----- 25
 Db 25499 PMSDKMYKKIKIDQYEMGEKLDREVQKRPKIRLSRMEQFVYMLPRTTDOYRPKWRIPK 25558
 QY 26 ----- 25
 Db 25559 LSQDDLEIVRPARRRTPSPDYFYRRPRRSGLISDEELLPIDYLA MKRTEERLRL 25618
 QY 26 ----- 25
 Db 25619 EEELEIGFSAPSPSPHFLSSLRYSSPOAHVVEETRNKFRSTYHIPTKASASTSY 25678
 QY 26 ----- 25
 Db 25679 AELRERHQAAYRQPRQORIMAREDEELLRPVTTTQHLSEYKSELDFMSKEEKSRRKS 25738

QY 26 ----- RPV----- 28
 Db 25739 RROREYTEITEIEEVEYISKHAQRESSSASRLRRRSLSPTYTELRPVSELIRSPQ 25798
 QY 29 ----- 28
 Db 25799 PAREYEDDTERRSPPPERTPRSPSPVSSERLSRFRSARDPDIERSYEMKALKTQKT 25858
 QY 29 ----- 28
 Db 25859 SERKEYVLSQOPEPTLDHARITLIMRSHRVPCGQNTFLVYOSKPTAEVKYHNGVELQ 25918
 QY 29 ----- 28
 Db 25919 ESSKIHVTNLSGVLTLEILDCHTDGTYRAVCTNYKGEASDYATLDVTGDTTYASQR 25978
 QY 29 ----- 28
 Db 25979 RDEEVRPSVPELRTREAVAVPSFKTSEMASSSVREYKSOMTETRESLSYEHASAE 26038
 QY 29 ----- 28
 Db 26039 MKSALBEKSLEKSTTRKIKITTLARILITKPRSMYVEGESARPSODTGEPPVTYWL 26098
 QY 29 ----- 28
 Db 26099 RKGQVLTSAHQVTTTKYKTFEISSVOASDEGANSYVENSEKQEAFTLTIOKARV 26158
 QY 29 ----- 28
 Db 26159 TEKAVTSPPRVKSPEBRVKSPEAVKSPKRVKSPEBHPRAVSPTEKTPPIEKVOHLVPS 26218
 QY 29 ----- 28
 Db 26219 APPKITQFLKAEASKELAKITCVESVLRKAEVYTKDKGKLKENGHPHYASDGYE 26278
 QY 29 ----- 28
 Db 26279 LKINNLTESDGEYVCEISGEGSTKTNLOFMQAFKSIHERVSKISETKKSDDKTTEST 26338
 QY 29 ----- 28
 Db 26339 VTRKTEPKAPEPISKRPIVYTGLODTVSSDSVAKFAVATGEPRTALWTDKRAIYOG 26398
 QY 29 ----- 28
 Db 26399 GKYYLSEDKGFLEIHKTDTSGLYTCTVKNSAGSVSSCKLTIKAIKDEAQKVSQ 26458
 QY 29 ----- 28
 Db 26459 KTSBITPQKRAVVOEISORALRSEIKSEAKSOEKLALKEBASVLISEEVKSAATS 26518
 QY 29 ----- 28
 Db 26519 LEKSIYHEIITKSQASEBVRTAEIKAPSTQMSINEGORLVKANIAGATDVKKWLVNGV 26578
 QY 29 ----- 28
 Db 26579 ELTNSEERYGVGSGDQTLTIKQASHRDEGILFCISKTEGIVKQCYDILSKELSDABA 26638
 QY 29 ----- 28
 Db 26639 FISOPRSQINEGONVLTREISGEPSEPEIEMFKNNLPISISSNVSISRSRNVYSLEIRN 26698
 QY 29 ----- 28
 Db 26699 ASVSDSGTYTIKAKNPRGQCSATASIMVLPJVEBPREVVLRTSGDTSLOGSFSOSQVOM 26758
 QY 29 ----- 28
 Db 26759 SASQOASFSFSSSSASMTMKKFASMSAQSMSMQESFVEMSSSSFMGISMNTQLESS 26818

QY 29 -----G 29
 Db 26819 TSKMLKAGIRIPKPIKALPSDISIDEKVLTVACAFTEPTPEVTWSCGGRKIHSGOEG 26878
 QY 30 RF 31
 Db 26879 RF 26880

RESULT 5
 T42629
 tenascin-X - bovine
 N:Alternate names: flexillin
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T42629
 R:Elfefflou, F.; Exposito, J.Y.; Garrone, R.; Lethias, C.
 J. Biol. Chem. 272, 22866-22874, 1997
 A:Title: Characterization of the bovine tenascin-X.
 A:Reference number: 222180; MID:97426436; PMID:9278449
 A:Accession: T42629
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-4135 <ELB>
 C:Cross-references: EMBL:Y11915; NID:g2462978; PIDN:CAI72671.1; PID:g2462979
 C:Genetics:
 A:Gene: TN-X
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin ty
 C:Keywords: extracellular matrix; glycoprotein; heptad repeat

Query Match 41.0%; Score 64; DB 2; Length 4135;
 Best local Similarity 0.4%; Pred. No. 2.4e+03;
 Matches 17; Conservative 5; Mismatches 8; Indels 4048; Gaps 4;

QY 1 SRHXH-----6
 Db 56 SRLYEHTVEGGEKQVFTHRINLPPSAGCCGPTPEPPVASEVQALRVRLLELLELYKG 115
 QY 7 -----6
 Db 116 LKEQCTGGCCPPAAQAGTGTDIRSLSLHGVDFSLKACSCPEWGGPTCSDEGAEGP 175
 QY 7 -----6
 Db 176 PSSPPSPAGSCPPDDCNDQGRVCYGRVCYCFPGYTGSPSCWPSGPDCHGRGRCVQGVYCVCR 235
 QY 7 -----6
 Db 236 SGFSGDCCSVRSCPRGCSQGRCEDEGRVCNPGYSGEDCVRSCPRDCCSQRGRCENGRCV 295
 QY 7 -----6
 Db 296 CNGEYTGDDCGVNRSCPRGCSQGRCEDEGRVCNPGYTGDDCGSRTCPMDGEGGRVCVDGR 355
 QY 7 -----6
 Db 356 CYCMPTGAGEDCSTRPCRCRRCRPRGCEDEGCICDEGISGDDCGVNRSCPSDCNQGRCEB 415
 QY 7 -----6
 Db 416 GRVCWMPYSGPDGACAPRDCRGRRCENGVCYCHAGYSGEDCVRSCPRDCCRRRGRC 475
 QY 7 -----6
 Db 476 ESRGRCVWPYTGRCDCGRACPDCCGRGRCVDCRCVCPNPFAGEDCGSRPCDGRGRG 535
 QY 7 -----6
 Db 536 RCGDGVSCDVGVEGEDCGRSGRCPRGCGRGQCLLEGRCVDDSYEGEDCGVRRCPRDCNQ 595
 QY 7 -----6
 Db 596 RGVCDDVCTCWBSGAFAGEDCGLRVCPSNCHRRGRRCENGRCVCDSGYTGPSCATRTCPADCC 655

QY 7 -----6
 Db 656 RGRGRCVQGVCHVWYSGEDCGEPPASACPGGCGPRLCSAGOCVCSGFRGPDCAI 715
 QY 7 -----6
 Db 716 QTCPCDRCRGRCRSGSCVQDQYAGEDCGEVPALIEGRMHLLETTVTEWTRAPGNV 775
 QY 7 -----6
 Db 776 DATEIPIPTTBGASPPFTARVPSSASAYDQGLAPGGEYQTVBALRGTNMGPPASKTI 835
 QY 7 -----6
 Db 836 TIMIDPQDLRVAVVPTTLELNLKPOAEVDRFVYSYAGNQVRLEVPSEADGTLIT 895
 QY 7 -----6
 Db 896 GLMPGEVYVTVTAERGRAVSYPASIRANTGSSLGLGATDEPPSGPSTTQGAQAPVL 955
 QY 7 -----6
 Db 956 QQRPOELAEVLGKDKTGRLRVAMTAQPDFTTHQLRLVPEGGAHEELLPGDVRQAL 1015
 Db 1016 VSPSPESPELRLGIPPSGEPAPLIYQIMDKDEKPKPLAPPLGLKILTVTDVTS 1075
 QY 7 -----6
 Db 1076 SLLEMTVEGEFDFSEVIOYKDRDRPQVVPVGPORSALISMLDYGRRKFKVLYGLVKK 1135
 QY 7 -----6
 Db 1136 RRGPLVAEAKILSOTDPSVTPPRLGNMTDPTDLSLHSLWTVDEGQFDFSMQYRDRA 1195
 QY 7 -----6
 Db 1196 GRPQVVPVGPDSYILSPLDPHKRYRFLFGIANKNRHPLTADGTAPEKKEPRHPE 1255
 QY 7 -----6
 Db 1256 PERPLLEGTLVAGATADSLRLSWTVAQSPDSFVQYKDAQGRQAVPVGDENEVAIP 1315
 QY 7 -----6
 Db 1316 SLEPDRKTKMNLXGLHGQRVGPVSVAITAPQEVLDETPSATEMETPSPTEPSTAPE 1375
 QY 7 -----6
 Db 1376 SPEKPLLEGELMTGSSPDSLISWTIPQGHFDSFTVQYRDGQDPRVMKVPDGEDGVYIS 1435
 QY 7 -----6
 Db 1436 GLEPDHKTKMNLXGFHDQRVGPMSVIGVTAEEETPSPTEMETPSPTEVEETPSMPER 1495
 QY 7 -----6
 Db 1496 STEAPEPAEPLLEGELTYTGSSPDSLISWTVPQGHFDSFTVQYKGRDQVAVYRGEBET 1555
 QY 7 -----6
 Db 1556 EVTIGLEPGRKTKMNLXGLHGQRSGFVSTVGTAPQAEETPATEPPEPRLEGELTVT 1615
 QY 7 -----6
 Db 1616 DVTPNVGLFMTVSGQFDSFVQYKDRDQSHVVPVADQREATVSGLEPERKRYRMNVY 1675
 QY 7 -----6
 Db 1676 GLHGGQVNGPLSVVALTPAPVPPDPVTEPPVPRLEGELTVTDVTDSDVGLSWTVABGERDS 1735

Qy	7	-----	12	-----	SMEXRT	12
Db	1736	FLVQKRDGQPOVVPVATDQREVTIPGLEPSRKKYKFLFGIODGKRSTVSVEAKTVGR	1795	1:1:1		
Qy	13	-----	12	-----		
Db	1796	GDASPGAPRRLGELMTVDPIDSLRLSWIVPBGHDSFVVOQFKORDGPRVVSVEGHERSV	1855			
Qy	13	-----	12	-----		
Db	1896	TISPLDSGRRYREFLYGLLGKRHRHGLTEGTETTRAVDAGTKRPSKPRLGELQVTG	1915			
Qy	13	-----	12	-----		
Db	1916	WTSNSVGLSWTPBGEHDSFVIOYKRDGQPOVVPVEGSRREVSYSGLDPARKYKLLYG	1975			
Qy	13	-----	12	-----		
Db	1976	LSRDKRVGISALAVTEBPAREEIKAEPAFSPAPASEPLLGVTLEAAPHSLRLSWTAT	2035			
Qy	13	-----	12	-----		
Db	2036	EGEEDSEFVQYTDENGQLQEVNNGDQHDITISLESDBHYLVSLYGFHDGQRVGAHIE	2095			
Qy	13	-----	12	-----		
Db	2096	AMTAPREDEDEPSLSSTQTPSTAVPERPHIKPRLGELAVDTTPDLSLSMTVPBGOFD	2155			
Qy	13	-----	12	-----		
Db	2156	HFLYQKNGDQPKVVRVPGDEDEVTISGLEPDHKYKMYGFHNRQMGVSVIGVTTA	2215			
Qy	13	-----	12	-----		
Db	2216	EERTPSPTMEMETPSPTMEMETSPTEBPNTVEPPEPEPFELELTVTGSSPDSLSTSVT	2275			
Qy	13	-----	12	-----		
Db	2276	POGHDSFTVOYKNGDQPKVVRVPGHDKGVITISGLEPDHKYKMYGFHNRQRVGPVS	2335			
Qy	13	-----	12	-----		
Db	2336	IGVTTAEETPSPTEVEETPSPTEPSTEAPPEPEPEVGLMVTGSSPDSLSTSVTPQG	2395			
Qy	13	-----	12	-----		
Db	2396	HFDSEFTVOYKRGDQPVVVRVGEETEVTVEGLGPHKYKMYGLHGRRVGPVSTAMT	2455			
Qy	13	-----	12	-----		
Db	2456	AREEPASPPLKPOLGELTVTDATPDLSLSMTVPBGOFDHFLVQYKNGDQPKAVRVP	2515			
Qy	13	-----	12	-----		
Db	2516	GDEDOVTISGLEPDHKYKMYGFHNRQRVGPVSTGLTVSEKQDQMTAPRTDLPTAPE	2575			
Qy	13	-----	12	-----		
Db	2576	PIKPRLGELVTDATPDLSLSMTVPBGOFDHFLYQKNGDQPKAVRVPGEDEVTIS	2635			
Qy	13	-----	12	-----		
Db	2636	GLEPDHKYKMYGFHNRQMGVSVIGVTTAEETPGTEMEQTPSPTEVEETPGTEV	2695			
Qy	13	-----	12	-----		
Db	2696	EETPSPTSTEAPPEPEPLGELTVTGSSPDSLSTSVTPQGHDSFTIOHKRGDPQ	2755			
Qy	13	-----	12	-----		
Db	2756	VVRNGEETEVTIGGLEPGRKYKMYGLHSGQRVGPVSTVGTVDQEVETPSPTPS	2815			
Qy	13	-----	12	-----		

Dd	2816	TEADPEEPBLDGLVTWGSSPDSLSLSWTVQGHFDSFTIQYKGRDRPQVYRVGGEE	2875
Qy	13	-----	12
Dd	2876	VTIGLEPRHKYKMNLYGLHGRRVGPSTVGVTAPDYDAMTQPNSTVDEPPTKRLG	2935
Qy	13	-----	12
Dd	2936	EVTYTDATPDSLSLSWTIPBGQDHFVQYKNGDQPAKAVRPGEDEGVTISGLEPDHKY	2995
Qy	13	-----	12
Dd	2996	KMNLYGFHDHQRVGPVSVIGVTTAEEBETSPTKMEETSPTEMEETSPTEBPEADBP	3055
Qy	13	-----	12
Dd	3056	EEPLLELVTWSSPDSLSLSWTVPOGHFDSFTVQYKGRDQVVRVGGEEETVVGLE	3115
Qy	13	-----	12
Dd	3116	PGRTYKMNLYGLHGRRVGPASTVGTASLTTERPLAPRLGELAVAVTSDTARLSMTVE	3175
Qy	13	-----	12
Dd	3176	QGPFDSELYQYKDVOGQPAVPAADLREYTVSSIAPGRKYKFLLEGLRDEKRRGPVSAD	3235
Qy	13	PDINPA----- 	18
Dd	3236	AKTLDPKRPAPRGLGELTVTDVTPGVSGLSWTVPEGEFDSFMVQYKDRDGPVVPVPAADQ	3295
Qy	19	-----	18
Dd	3296	REYTVPGLEPRKRYKFLLYLGRKRLGPIAEGSTAPLEKERQPPRLGELTVTDEPN	3355
Qy	19	-----	18
Dd	3356	SLRLSWTVAOGRFDSFVQYRGTDGPRLMVPAADQREFTVEGLEPRKRYKFLLYGLLG	3415
Qy	19	-----	18
Dd	3416	QRLGPASVIGMTAPEEDTPAPMHAATEAPKPPGPRGLGVLAVRDVSPDSLRLMSVQGP	3475
Qy	19	-----	18
Dd	3476	FDSFVQYODTDGQPAALVGGDNKVLSGLEPSTSYEFLLYGLHEGRKLGVPVSAEGT	3535
Qy	19	-----	18
Dd	3536	GPVPAGQTPGEPGRLSHLSVTDVTTISLRLNMEAPPEAFDSFLRFGVPSPSTILEPQR	3595
Qy	19	-----	18
Dd	3596	PLLOBELTVPCTRRSAVLRDLHGLPGLTYTLTYLGLGRPKRADSIOGNARTLSPVLESPDL	3655
Qy	19	-----	18
Dd	3656	QFSEIRETSARVSWTPPTSRVDGFKVSYOLADGGEQSVQVDGRTQKLEGLPGAOYEVT	3715
Qy	19	-----	18
Dd	3716	VVSVNGFESEBPLTGFLITYVPDGTJHRLALNLDESALLHMKRPQTPVDTYDVYKVTAPGA	3775
Qy	19	-----	18
Dd	3776	PSLQASAPGSAVDYPLQGLVHTNYATLRLGLRGNPNTSPASITFTTGLEAPOLEAKEV	3835
Qy	19	-----	18
Dd	3836	TPRATLLTWTAPEVSPPTGYLLSFNTPEGQTOETLLDPGVTSHQKGLFPSTPYSTWLRAM	3895
Qy	19	-----	18

D	b	703	AERAPWMDPFWHMRPHFDLLQVRBARLUNDHDDHDIETLPETLDDSENAPVAKLLAAYP	762
O	y	5	-----	4
D	b	763	NAKTKVNTRDEAMFPPLLIRKHVPMFWTTAIDGDKEMFAKDTLLMQADPRYDAGVRI	822
O	y	5	-----	4
D	b	823	IPGPVSAGITKKNEPVANLGRFEADTTALNDAGVAPVELYSRLSAKNAEFLRNAP	882
O	y	5	-----	4
D	b	883	TIMHGHILIANPAYELEBEAFDIYDGEFAIRINDSYRDNLPEQRPFYKHYDIPYA	942
O	y	5	-----	4
D	b	943	LSEAVATGASVVDARLPKAVPDLLAGVAGVSISETGDKITTELPRKIEGVSEENPYG	1002
O	y	5	-----	4
D	b	1003	LYEVSFTLPSTLLTAHTAVTGALGTANAGTPDALVPCMPAIYTLTGTGRLEEHGEPA	1062
O	y	5	-----	4
D	b	1063	GTDPPEVTEGLINAVHLDHVVYDVRVPLHELAKGEGGGRIDYTSRCASIENSNGRIYTV	1122
O	y	5	-----	4
D	b	1123	ELELMDATQEVVATQOMQORAINRGATGTSVPASAPSMGGKSGDKIETTPRSFVDRAIV	1182
O	y	5	-----	4
D	b	1183	TAPSDMTPFALVSGDYNPIHTSTNAARVNLDAPLVHGMLSATQHLAGNHGVGWYTY	1242
O	y	5	-----	4
D	b	1243	SMYGMVOLNDEVELTVERVGRKGIHAFFYTCXIDGEVYSRQOALMAOPRIAYVYPGOGI	1302
O	y	5	-----	4
D	b	1303	QAEQMGGRDSDASAAAEVWRRADRHTPTALGFSIRQIIDDNPTELVRGTFFVHPNGVL	1362
O	y	5	-----XMSM-----	8
D	b	1363	HLTPTQVALAVAYIAQOTERLREADALGTSMSKAGHSIGETYLALASLANIFDLEAVIDIY	1422
O	y	9	-----	8
D	b	1423	YSRGSAMGLIVERDENGSNMYGKALRPMNIGVPADQVEAYIAQTAHEETGELEIVANNI	1482
O	y	9	-----	8
D	b	1483	AGOOYSTAGTKAGLAILKKKANSVKDRAIVTVPGIDVPHSQVLRDGVAPAEKIDELLP	1542
O	y	9	-----	8
D	b	1543	ETLDDALVGRYVNVNLVLPFELTQEFVDKVYKPLAPSGKLDMLKVEDIDEQAPSRLMTE	1602
O	y	9	-----	8
D	b	1603	LLSWQFASPVRIETQOULLFEEDVQITIEVGLASSPTLITLMAKRSMDIAGVDLPVFNVERD	1662
O	y	9	-----	8
D	b	1663	QDQVMLDQVGEAPASFDVBEGBATSTAASETPGBSAAASDNTQALPSAPOTVAEAP	1722
O	y	9	-----	8
D	b	1723	APSAPAGSTRADADBDLPFTAALMVLFAFONKIRODOINDSDVEELTNVSSRRNQ	1782
O	y	9	-----EXRTP-----	13

Db 1783 LMDMSAENRVPALIDGAADAVATLRRVKTAAAGSPGTVLSEAITARLQLTGAGV 1842
 QY 14 ----- 13
 Db 1843 KPAVISERTGTWGLPMSMAHAHEAELLGSRDESDVSGSLSTVPSAASKADVALVD 1902
 QY 14 ----- 13
 Db 1903 AAYQVAAAAGTSVSHGAASGAGGVVDSALDAVADIVTGENGLATPAARQVLAQGL 1962
 QY 14 ----- 13
 Db 1963 VEEAPETPETDNTLFAATEALGSGMEKVTVPSEDAKRAVLFDNRWASAREDLARVALGE 2022
 QY 14 ----- 13
 Db 2023 IDLPVKRFQGTGETIAKQAEWMAENTASTGAHAKATAETLHAIAAAAAAREELDGEFAGD 2082
 QY 14 ----- 13
 Db 2083 VALVTGAAPGSIATFALVERLLEGATVYIMASHVSQSRKEFAKLYAAHAIPGALMWYP 2142
 QY 14 ----- 13
 Db 2143 ANLRSYRDVALIDWIGNBORASVQNEVKITKPALEPTLAPEPAASVSGSVADAGPOAE 2202
 QY 14 ----- 13
 Db 2203 NOTLLMSVERTIAGLSNAQGVDTKCHIVLPSPNRMFGDGAHYGVKAALDAITA 2262
 QY 14 ----- 13
 Db 2263 KMSAAGMPGVTLLAQAKIGWSGTSLMGNDVILPAAEAAGIHWDPPEISSQLISLAS 2322
 QY 14 ----- 13
 Db 2323 EESRAKAAEAPLELDLIGLGSSNISISELAQAREDAEAQASGDADAAPAAATIP 2382
 QY 14 ----- 13
 Db 2383 ALPRTREVELPALPEGEVDVTTDDDMVYIAGVGVSSWGSGRTRFEAEYGLQBDGAV 2442
 QY 14 ----- 24
 Db 2443 DLTAAGVLELAWMTGLISMSNDPRAMYDEGEVDEADIIYARDEVARSGIRLTLDK 2502
 QY 25 ----- 24
 Db 2503 YNMVDOGSIDLTSVFLDRDVIYFTVPTQEBALDIEADPSFTKLREVDGEWEVTRLAGATA 2562
 QY 25 ----- 24
 Db 2563 RVPKATLITRTVAGQMPDHFDAAKMGIRPHMLDALDMVAMNLTVAVDAFTQAGFPAEL 2622
 QY 25 ----- 24
 Db 2623 LQVHPAQVATTOGTGIGMESLHKVFVTRLLGEDRPSDILQELPNVIAAHTMQSLVGG 2682
 QY 25 ----- 29
 Db 2683 YGSMIHPIGACATAAASIEEGVDKIALGKADLVVAGSIDVQVESLTGFQDMNAATATTK 2742
 QY 30 ----- 31
 Db 2743 MTDGIDDRF 2752

RESULT 7
 T03222
 probable polyketide synthase module 2 - Streptomyces hygroscopicus
 C:Species: Streptomyces hygroscopicus
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 03-Nov-2000
 C:Accession: T03222

R:Ruan, X.; Stassi, D.; Iax, S.; Katz, L.
 Gene 203, 1-9, 1997
 A:Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus AT
 A:Reference number: Z14848; MUID:98085969; PMID:9426000
 A:Accession: T03222
 A:Status: Preliminary; translated from GH/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1762 <RUA>
 A:Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AC38062.1; PID:g2624949
 C:Experimental source: ATCC 29253
 C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-ox
 homology; [acyl-carrier-protein] S-malonyltransferase homology
 C:Keywords: [acyl-carrier-protein] S-malonyltransferase homology
 F:54-454/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
 F:550-822/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
 F:1598-1669/Domain: acyl carrier protein homology <ACP>

Query Match 39.78; Score 62; DB 2; Length 1762;

Best local similarity 2.0%; Pred. No. 7.6e+02; Mismatches 10; Indels 713; Gaps 3;

Matches 15; Conservative 4; Mismatches 10; Indels 713; Gaps 3;

2 RXH-----XSMEXR----- 11

Db 834 RNHNVEYRSTNEALGRVSGVSHVDKRALTPGKRTSLPTFPQRDRVLDHDAAGGAIVE 893

QY 12 ----- 11

Db 894 GAGLGTDHPLLGAVTVADTGELISGLSTSHPMULTDHTVNGTVIVPGLALDLAH 953

QY 12 ----- 11

Db 954 AAEHTDHTVDELVIHTPLALHTPPSSQTVGAETDGNRPALHSRDTGTWRHTTGT 1013

QY 12 ----- 11

Db 1014 LSNQTHPAALSTNSPTAHQIDLTAYQQLADTGLHGPAGGHLTYQDRNTLFAEIE 1073

QY 12 ----- 11

Db 1074 LPEAGLPTGILHPALLDAALHPAASARDESGIQLPESWAGVTWSTGATHLRVL 1133

QY 12 ----- 11

Db 1134 DYDDGAIKRLATWDTGQPVITIDSLTRPLHPDQAAOPEYNHLYTLGMSPPPTPTT 1193

QY 12 ----- 11

Db 1194 TGHFERVTPAGDDLLADTNRITQAOTLRIOQHIAADTTPIVVEATHDDLAGARWGL 1253

QY 12 ----- 11

Db 1254 IRTAQTEHPGRILLIDTDNHPTSRDALASVAGKPAIRGATITRLRLSVDRAS 1313

QY 12 ----- 11

Db 1314 TERRPDSMGVPLVTGGTGLSLIARHLVTEHGVLDLVYKSGPADGATELTNQLQ 1373

QY 12 ----- 16

Db 1374 HGARIRITISCDLTDRFALTLTLDLDELGPLGVHTGALADTTIDHLNPDLTTLTPKAN 1433

QY 17 PAMY----- 20

Db 1434 PAWMLHELTQDDHLALFVSSVAGVLNAGAGANYAANSFLDALITRRRQGLPGTSLA 1493

QY 21 ----- 20

Db 1494 WGMWOREGGMTAHLTQADHQRMTFAGIHGLDABGTTLPDTALDRGLCAVAPVKILHPTL 1553

QY 21 ----- 30

Db 1554 NRADTVPAVLRLGVAVRPAAR 1575

RESULT 8
 T34022
 zonadhesin - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
 C:Accession: T34022
 R:Hardy D.M.; Garbers, D.L.
 J. Biol. Chem. 270, 26025-26028, 1995
 A:Title: A sperm membrane protein that binds in a species-specific manner to the egg ext
 A:Reference number: 221464; MUID:96064658; PMID:7592795
 A:Accession: T34022
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-2476 <HAR>
 A:Cross-references: EMBL:U40024; NID:g1066465; PID:g1066466; PIDN:AAC48486.1
 A:Experimental source: strain Meishan; testis
 C:Genetics:
 A:Gene: zan
 C:Function:
 A:Description: may be involved in sperm adhesion to the zona pellucida

Query Match 39.7%; Score 62; DB 2; Length 2476;
 Best Local Similarity 1.6%; Pred. No. 1.5e+03;
 Matches 19; Conservative 2; Mismatches 7; Indels 1133; Gaps 5;
 QY 4 HXH-----SHEKRT----- 12
 | | | | |
 Db 762 HCHCRPSSRMCCQTFKCGTHTVCQKNGQYCHPYGSATCYGDPHYLTFFDGRNFMG 821
 QY 13 ----- 12
 Db 822 KCTIILAOPCGNLTENHFFRVLVKKERQEGVSCLSKVYTLPESTVTLKGRHTLVGGQ 881
 QY 13 ----- 12
 Db 882 RVLPAIPRSQVFLAPSGRVEVLTAFGLRVNMDGQQLFVSPSTFSGKLCIGCDYDG 941
 QY 13 ----- 12
 Db 942 DSSNDNKPDPGSPAKDEKELSSWQTSSEADQOCENQVSPSCNTALQNTMSGPERCG 1001
 QY 13 ----- 12
 Db 1002 LVAPHGVFEACPLHLRASSFFKSGTSPMCMFQGLQHLCAHMSALTENCODAGYTVKPMR 1061
 QY 13 ----- 12
 Db 1062 GPQFCPLACPRNSRYTLCARLCPDYCHSEFSGRACKDRVCEGCEDPGFVLSGLQCVSR 1121
 QY 13 ----- 12
 Db 1122 ECGCLDSTAGYKVGKRWKFRPGRCGLICBGNRTKCVLMRCQAQEFCCGQDGIYCHQAQ 1181
 QY 13 ----- 12
 Db 1182 GSATCYVSGDPHLTFDGLAHHTGTCTTTLTKPCWLRSLNSFLYSATNERGMLNAS 1241
 QY 13 ----- 12
 Db 1242 YRAVQVQVENLRSLIKGRKVTLDGRVVALPLMPOAGRVSTISSGFFILTYDFGLQVR 1301
 QY 13 ----- PDINPA----- 18
 Db 1302 YDGDHLVEVTVPSSVAGRLCGLCGNTNNSLDDIILPDKRPASSSVRLGASWKINELSEP 1361
 QY 19 ----- 18
 Db 1362 GCFABGGKPPRCIGKEVADAMRKNCVLMNPGPFSCCHRVYAPSSSSCLYGQCATKG 1421
 QY 19 ----- 18
 Db 1422 DTLTLCRLQAVASLCARAGQALTMWNGTFCPLKCPSSSSYSTCANPCPATCLSLNNPSY 1481

QY 19 -----WYX----- 21
 Db 1482 CPSTLPACBECBQKXHIISGTSVPLSQCCTGTQSGYHPVGESEYTDNCSRLCTGSA 1541
 QY 22 ----- 21
 Db 1542 HNNISCRQASCKPQCMQFQDGLIRCVAGMGVCRIPTDTSHTVSPDGSYHAYRGNCTVYL 1601
 QY 22 ----- 21
 Db 1602 VKICHTMDLPFKISGENKREGQPPAPVLYRQVYVDIFNTLVTLKQDQVLINGTRVSLP 1661
 QY 22 -----XRGIR----- 26
 || | | |
 Db 1662 ATQINGVAVISNDGYTLINIGVQVKNPDGGELEVEIPKAYGRGCGVCSNFDEED 1721
 QY 27 ----- 26
 Db 1722 ELMPSPDALLDVNYVDSWRDKEIDPNCQEDDRKTEARSQPSANCRPADLERAQRC 1781
 QY 27 ----- 26
 Db 1782 QAAFAQAPANACATRVYLSPIVRSCTHKLCEFGGLNRAFCESLQAFGAQAQGIKPPVM 1841
 QY 27 ----- 26
 Db 1842 RNSSPCLDCSAHSVYTSVPSCLSPSCQDPBEGQCTGAGAPSTCEGCTCEPGYVLSQOC 1901
 QY 27 -----PVGRF 31
 || | | |
 Db 1902 VARSQCGCRDANGTFLPVGRF 1922

RESULT 9
 T14591
 actinomycin synthetase II acmb [imported] - Streptomyces chrysomallus
 C:Species: Streptomyces chrysomallus
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 03-Nov-2000
 C:Accession: T14591
 R:Schauwecker, F.; Pfennig, F.; Schroeder, W.; Keller, U.
 J. Bacteriol. 180, 2468-2474, 1998
 A:Title: Molecular cloning of the actinomycin synthetase gene cluster from Streptomyces
 A:Reference number: 218152; MUID:98233744; PMID:9573200
 A:Accession: T14591
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-2611 <SCH>
 A:Cross-references: EMBL:AF047717; NID:g3114611; PID:g3114612; PIDN:AAC38442.1
 C:Genetics:
 A:Gene: acmb
 C:Superfamily: peptide synthetase ppsd; acetate-CoA ligase homology; acyl carrier pro
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
 F:509-966/Domain: acetate-CoA ligase homology <ACLI>
 F:982-1050/Domain: acyl carrier protein homology <ACPI>
 F:1578-2015/Domain: acetate-CoA ligase homology <ACLI2>
 F:2031-2098/Domain: acyl carrier protein homology <ACP2>
 F:1014-2063/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 39.7%; Score 62; DB 2; Length 2611;
 Best Local Similarity 0.8%; Pred. No. 1.6e+03;
 Matches 19; Conservative 5; Mismatches 7; Indels 2346; Gaps 6;
 QY 1 SRX----- 4
 | | | | |
 Db 116 AREHLFDHALLRLGDDRCYWGTHHYMDAFGYVLTTRTAEYVTTALABRPVSPSPFG 175
 QY 5 ----- 4
 Db 176 SLDGFIADQKRYDSEGTGDRAYWTERFADRPETGIVARSTTPHYLRTAALPGE 235
 QY 5 ----- 4

Db 236 LDGLREARAGBARAPWSHLIAAVVYLHRMTGATDVVILGPTARLDQLOQRTTGTASNY 295
 QY 5 ----- 4
 Db 296 LPLRLTVRPELTLKOLLTOVSARIVELGRHORRYRAEDLQDRLDPLGGLTWPAPVNNIMS 355
 QY 5 -----XHS----- 8
 Db 356 FDATAFTAGLPTTAHSLSSGLVGDLTFAANDRDGAGLTVDLNAHPELCEPERELTAHGR 415
 QY 9 ----- 8
 Db 416 LVATLRAIASDGRIDRPGRIDLLTAERRALLAGPEDAPATTGATTEATLPELFRATAA 475
 QY 9 ----- 8
 Db 476 APDAVAVVCDSTLTYELDERANRLAHLAAGVGPRIVALALPRSDLVVAVALAVLK 535
 QY 9 -----EXRTPD----- 14
 Db 536 AGAAYPLDPEYRANRLAHWTDAQPTLLITTTETAKLPDRHTAPALRLDDETLAALA 595
 QY 15 ----- 14
 Db 596 GOPANSPAVGLRDPHPAYVITSGTGVPKGVNTHRNVRLEPDATRPWDFGPDVWTL 655
 QY 15 ----- 14
 Db 656 FHSTAFDESWELMGALLHGRVYVPRYDVSRSFHAFLDLADOGVTVLNOTSAPHOA 715
 QY 15 -----INPA-----WT----- 20
 Db 716 QAAADPGRPRLALRTVYFEGEALQPARLAEMVRRHREDTPQVNMVGTETTVHTQ 775
 QY 21 ----- 20
 Db 776 PLTRDRAAGAASVIGASISDLRTHVLDDGLQVPGAVGELYVAGLARGYLGRPALT 835
 QY 21 ----- 20
 Db 836 AERVADPYGARGMYRTGDLVRNPDGELEFVGRADHOKVAGFRIELGEVEBAALLAH 895
 QY 21 ----- 20
 Db 896 PDVEQATVYREDRPGDRLVAVVGREALRPEQVRETRERLPEHNVPAVVOLERLPL 955
 QY 21 ----- 20
 Db 956 TPNGKIDRALPEPDFALAGTGREARTPOEQIVCDLFTOVLGLPRVGVDDFEELGSHL 1015
 QY 21 ----- 20
 Db 1016 LATRLIAHLRTVGLVELELRSLFBGPTPAVAARLDTAGPQRLALTVRQRPVWPLSSAQ 1075
 QY 21 ----- 20
 Db 1076 RRLMFLSTLEGPATYNIPIVLRISGRIDVPALCALGDVGRHESLRTVEPVDTPTQ 1135
 QY 21 ----- 20
 Db 1136 RVLPEAARPLTVTPISEADLPNALAKAGARAFDLAGEPPLRTLELSELSREHVLVVA 1195
 QY 21 ----- 20
 Db 1196 HHIAADGSMERPLSRELEAARAEGRAPQWAPLPVQYADYTIQNELLDQNDPDLF 1255
 QY 21 ----- 20
 Db 1256 ATQVAVYMETLAGLPDQITLPTDRPRAVMTYRGDYLVDDIDPELHRLTELANGSASL 1315
 QY 21 ----- 20
 Db 1316 FMTVLAGLAILKRLGAGDDIPLGSPJAGRTDQALDLDLIGFTVNTLVLRIDTTGDPPTFTQ 1375

QY 21 ----- 20
 Db 1376 LTRVRENSLAAYAHQDVPFEVILEVLANPRTIAHPLFOQIMIALQNAPENEFSLPGIRA 1435
 QY 21 ----- 20
 Db 1436 GIELGRTGAKFDLFFSLTERGPDGEPQGLTGEVESSDLEAATVQALFDRVRLHA 1495
 QY 21 ----- 20
 Db 1496 VTAQPEQPLSRIDVLTPEERNFTIVEVNRTELPLPDASIAELFEQVTLTPDAPALVSDG 1555
 QY 21 -----XRGIRP----- 27
 Db 1556 ATLSYSELNTRANHLAHOULTTGRPGDAVANLLQSPDVTYVALAKTGATYIPIDSR 1615
 QY 28 ----- 27
 Db 1616 YPADRIRLVLDETRKLITDHTTDLDTYTTQFNPADTPHDGEDGPNHTHPDAAYI 1675
 QY 28 ----- 27
 Db 1676 MTTSGSTGRPKGVIAITHNITLALDPRDPTAHRVLLHSPFADASTYEIWPPLNGN 1735
 QY 28 ----- 27
 Db 1736 TVVLAPTDGDVHTYHRVITDQITAVELTALFNLTEHDPACLAGVREWVTGEAVSA 1795
 QY 28 ----- 27
 Db 1796 FSVRVQAEPSVVVDVYGPETTTTFAHNPPTPYTGPAVVALGRPMATMAVYLLDA 1855
 QY 28 ----- 27
 Db 1856 LQPVAGVVGELYLGAGLAGYLDRPALTAERFANRHRGEMRTFGDLARMSADAOL 1915
 QY 28 ----- 27
 Db 1916 EFVGRADQOVKRGFRIBEPGEIENVLTGHPAVAQAAILVREDQGRPRVLAVVADGSTA 1975
 QY 28 ----- 27
 Db 1976 PDGLREALRRLLPEYWPSTAFVYLDRLPTLANGVDRALPAPEVSRAVSAREPTREG 2035
 QY 28 ----- 27
 Db 2036 ILGDLEAQLGYERVGLEYDFALGDSIVSTRVSLARSAGIGSVNDVFEQRTAAGLA 2095
 QY 28 ----- 27
 Db 2096 VVAEELTASDGEADESAGVPTPTIMRFDERGSDIRFHQAMLLQVPAALGERLVTYA 2155
 QY 28 ----- 27
 Db 2156 VOALLDHDHALRTRVPSDEGDGPEWELHIPAGTISAADLVRRVLDGPDDEATVTRHL 2215
 QY 28 ----- 27
 Db 2216 DAAQARLHAERGVLLQAVVFDAGPRRGRLLLVNHLVVDGVSWMILLPDLVAAMEATAQ 2275
 QY 28 ----- 27
 Db 2276 GDPRLPEVPTSPGRWSRLIAEARBARABAAALWTEVLTPVDEPLNATRPRDRTDTVG 2335
 QY 28 ----- 27
 Db 2336 RASLTVSLPEEVVSPLLTTPAFAHGVNDVLLALALAVARMRGRGGRGPHALLVDV 2395
 QY 28 ----- 27
 Db 2396 EHGREGREIYDGTLSRTVGMTSLIYPRLDPGALGWEVETGAPALGPKLRKVEQLREL 2455

QY 28 -----VGRF 31
 Db 2456 PDRGLGGLRLHRLAPATGPVLAGLAPOLGPNLIGRP 2492
 RESULT 10
 A35548
 319K protein ndvb - Rhizobium meliloti
 C:Species: Rhizobium meliloti
 C>Date: 21-Sep-1990 #sequence_revision 21-Sep-1990 #text_change 08-Oct-1999
 C:Accession: A35548
 R:IdP, L.; Dylan, T.; Dillet, G.S.; Helinski, D.R.; Stanfield, S.W.
 J. Biol. Chem. 265, 2843-2851, 1990
 A:Title: The ndvb locus of Rhizobium meliloti encodes a 319-kDa protein involved in the
 A:Reference number: A35548; MIMD:90153914; PMID:2154461
 A:Accession: A35548
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2870 <IE>
 A:Cross-references: GB:J05219; NID:9152270; PIDN:AAA6305.1; PID:9152271
 C:Keywords: transmembrane protein

Query Match 39.7%; Score 62; DB 2; Length 2870;
 Best Local Similarity 0.8%; Pred. No. 1.9e+03;
 Matches 14; Conservative 4; Mismatches 10; Indels 1694; Gaps 3;

QY 4 HXHSMEK----- 11
 Db 180 HTSTVRESITAMVEGFEHETLKIGELMALPSILRFLIENIRRAIRVERSGMRK 239
 QY 12 ----- 11
 Db 240 ANEVADQLRLNDEPGRTLLVESEALADNTITAOILYMRDSSGCAVIAWIEERLE 299
 QY 12 ----- 11
 Db 300 RRGDVEEALVAQNRLSGNATMNIIRSLREIDDTMAWVESYSKIDATLREGSDYA 359
 QY 12 ----- 11
 Db 360 ALDPSRNYRDTLEKLARRSGHEHEVTEIALEWVEAKAAVAEAPLOEPNVSFLVG 419
 QY 12 ----- 11
 Db 420 KORALEKRIGYSPSIFQHLIRSVRKLDWEALAGPNILITIAMIVYAVSPMDIPSCA 479
 QY 12 ----- 11
 Db 480 KLIMLLFALPASEGAMGLENTVETLFAPSRVLGYEFLDGPEDARTLVVPCLIARND 539
 QY 12 ----- 11
 Db 540 HDELYRNLEVHYLANPGEIYFALLSDWADSKSEAPADTVLEYAKREIASLSARAY 599
 QY 12 ----- 11
 Db 600 DGKTRFELHRRRLYNDEAGVWGMWKRKRGKLEHMLNLKRGDRDTSFLOGANNVPSGYQ 659
 QY 12 -----TPDIPA----- 18
 Db 660 VMTLSDTFLMRDAVTKLVGKLYHPINRPVYNERTQEVVYGSLLQPRVTPSLTGSSEAS 719
 QY 19 ----- 18
 Db 720 AFRIFTINKIDPYFYVSDVYQDIAGEGSTFGKGLYHVDFAEALAKRIEENAVLSHD 779
 QY 19 ----- 18
 Db 780 LLEGSAKCALVTDIELVEDEPIRYEVEMSRQHARAGDWQLLPYIFNPKNGLSMLGRMK 839
 QY 19 ----- 18
 Db 840 MYDNLRSLLIPVAMLAASVGMWYMEPTPALIMQVLIFSFLVAPTLISGISIMPRRNDI 899

QY 19 ----- 18
 Db 900 VARAHLFTVLSDIRANAQVALRIEIAHNNAMADATVRSIYRTFVSKIMLEWRTAAQ 959
 QY 19 ----- 18
 Db 960 VOSAGHSIGDYFRAMMTAPALALVSLALMAISDTGLPFIQLPALIWAASPAVAMEVSQ 1019
 QY 19 ----- 18
 Db 1020 SAETEDQVYSEALIEEMKRIARTWRFEAVTAQNFPLPDNFQETPPQVLAERTSPT 1079
 QY 19 -----WYXRGIRP----- 27
 Db 1080 NIGVYLLSVMSARSPGWIGFEETITRLQTIATIDRMPKYGHLNMYRTRGLEPMERY 1139
 QY 28 ----- 27
 Db 1140 VSSVDSGNLGHILAVSSMKREMAEAPSAHVQGNLDIGVAAIILKALNELPDDRKTVR 1199
 QY 28 ----- 27
 Db 1200 PLRLVEERLAFQNALAAVKKREREPASTRVINLAVLARMDHKLTVNLDEHYRTVQSGEV 1259
 QY 28 ----- 27
 Db 1260 ATWAGSLVACBAHTADGVFDGAIETALRQRLVLEKARDIAFSMDSPFLRPERRLLS 1319
 QY 28 ----- 27
 Db 1320 TGYRVANLEBACYDLASERLTSPLAIAGDLPTHEWYKLGRIPIVIGARGALVWS 1379
 QY 28 ----- 27
 Db 1380 GSMFEYLMPLVMOERQGLNQTNNLVYQDINNGRRLGTPWGISAEAFNARDHELYQ 1439
 QY 28 ----- 27
 Db 1440 YNFGVPTLGLKRGONAVIAPYASIIACWDPKALANLARLEVGALGAYGHDAVD 1499
 QY 28 ----- 27
 Db 1500 FETPRVPSQKCAVVRNTYAHHGHMSVAANVFNQGLREMPHADPYEAEALLQEK 1559
 QY 28 ----- 27
 Db 1560 PRDIPVMAKREPEALGKQADLREPEVRYEDPINDRETVLSNGHYSMLTATGAGY 1619
 QY 28 ----- 27
 Db 1620 ARMGQSVTRWTPDEVEDRTGTFIFLRDVTGDMWSATAEBRPAEGEKTVRFGDKAEF 1679
 QY 28 ----- 27
 Db 1680 VKTVDLTSVEVCIVATEHDAAGRVILLNTGTEDRFLEVTSYAEVYLAADDADSSHPTE 1739
 QY 28 ----- 27
 Db 1740 SKMFLRETSRHDVYIWSRNKRSPGDDPIEVAHLVTDNAGSERHTQETDRRRFLGGR 1799
 QY 28 ----- 27
 Db 1800 TLAFAAFDPGATLSGTGDTLDPVLSLRVVRVPAGKVSIVFWTIAAPDREGVDRAID 1859
 QY 28 -----VGRF 31
 Db 1860 RYRHPETFNHHLIAWTRSYQVMKRVGTISKEASFOMLGRT 1901

RESULT 11
 T12117
 polyprotein - fava bean dsRNA replicon

C:Species: *Vicia faba* (fava bean)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: F12117
 R:Referrer: P. 79, 2349-2358, 1998
 J:Gen. Virol. 79, 2349-2358, 1998
 A:Title: Nucleotide sequence, genetic organization and expression strategy of the double
 A:Reference number: 217424; MUID:98451319; PMID:9780039
 A:Accession: F12117
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-5825 <PEP>
 A:Cross-references: EMBL:AJ000929; NID:93184155; PIDN:CAA04392.1; PID:93184156
 A:Experimental source: Viroin; cultivar 447
 C:Comment: This gene product may be cleaved into several proteins including helicase and
 C:Genetics:
 A:Genome: dsRNA replicon
 C:Superfamily: fava bean dsRNA replicon polyprotein

Query Match 39.7%; Score 62; DB 2; Length 5825;
 Best Local Similarity 0.4%; Pred. No. 7.4e+03;
 Matches 17; Conservative 1; Mismatches 10; Indels 4154; Gaps 4;

```

OY 4 HXHSXEXT----- 12
Db 1336 HAHESEAGTSCCLLPAGACPCCGVNRQIESEILFENSEESADEAEONRYRKNSKP 1395
OY 13 ----- 12
Db 1396 RRPDEKGNRSNNRNDNRDANHHNSVNYHHGHNKQROGATQOPPKRSALHPEND 1455
OY 13 ----- 12
Db 1456 NDTPDSTPTVILLADPTNPGERQAPGDNIPEHNDIPGSSSTTQSSPPDDTNYSGPE 1515
OY 13 ----- 12
Db 1516 HNRNIONHRIRLIAQTFNEPFGDLAYGNDIAVKKSHKMLPPNORLRSCHFQFNMGDSI 1575
OY 13 ----- 12
Db 1576 YATENLEVIKTYNTSGEGGYGNALKILYPNLDLLEMQEIYGSSTQFSDWEIMRYAQ 1635
OY 13 ----- 12
Db 1636 AKOLNLIWTERCALVNSVCSNEGVICHGRHGMLEHWEALALOKGFADYHPTFN 1695
OY 13 ----- 17
Db 1696 ALTRBDLLDPAKSSGLKNNKYSFVGLDPRLLIKLOTBMHESLSKYVEDVNPAGFVKITKG 1755
OY 18 ----- 17
Db 1756 STHYVTNDRNNLGYSPQGTFTAQTTEDNQELVELLLAAYSAPARVITDWFYRPESP 1815
OY 18 ----- 17
Db 1816 ADATGIDHYGKQIVDAEVNOACAGDLIKSELIONRVDCKEYINNHLKVKOTGRLK 1875
OY 18 ----- 17
Db 1876 PDDLISGLIVGIQSTSYDSHVIGIVITNOIMCTCYTEKCAKATIKIVDLRKNTGSL 1935
OY 18 ----- 17
Db 1936 RALFGIFRMDFRDITLLEKATAVDALAGMGKSTPEIKLVNODCTVAQTSAAVSNILE 1995
OY 18 ----- 17
Db 1996 KLEGGKRNCKVMSIEKMTQOVNTPFLVDEASMTWETLSLTGPQVENVLYGNTL 2055
OY 18 ----- 17
Db 2056 QICVLDWTRTGGSRATKSLIEQAGIIRRHYYTHRIGNPLARELSLVYKELTTNAKHETNF 2115

```

```

OY 18 ----- 17
Db 2116 CTKSMDAVRMAELTSLAGSLIEPVILCFYNNVRAVMNYLKVCGRVDTIHKFQLEADNV 2175
OY 18 ----- 17
Db 2176 IYLWCPTGQTPGRITLDRHQCLSAATRAKMLWISVNEYSNNVPLHRKMGATTGSKH 2235
OY 18 ----- 17
Db 2236 TOPDTENNIADKTLQVYSHLSKVMITYSKDKITERLECSRLOSQEBGNLYMLKNTLTHLD 2295
OY 18 ----- 17
Db 2296 VDYVSGNDQSLRQSGSKLNEKFSMLSLFSKEPAPDVHSSSTLGKLAIVDQANAND 2355
OY 18 ----- 17
Db 2356 RINSQIMNIYTVQKNLAISNGSIAAAYTPVLTAMLQRPITKAECSPDEPTETLTANI 2415
OY 18 ----- 17
Db 2416 LMLGVNVLKIQDLFRNDKVIISGGILRSFYADSSSEIKQLYNOKFNCCGHVYLKYNPDTE 2475
OY 18 ----- 17
Db 2476 ISVKKKMFERYVLMIYVFGVCLSGSLHSVTIENELKMQSYGSGSLCGMVFHHNK 2535
OY 18 ----- 17
Db 2536 PIIVGDSYNTGNSRTVRENSKQDKGPIIAYLRMLNVDSGTVKLPYDPVGRGDEDM 2595
OY 18 ----- 17
Db 2596 FEGDNMAALVIERVLYKGPMAAALAKMLTNTNQLCNFYDNNSELPDSIKARLESNG 2655
OY 18 ----- 17
Db 2656 FTNFTNLVNEFVGTRENGFTNTELVNVSMDLCRPICTLDKRYLALMDQDN 2715
OY 18 ----- 17
Db 2716 LIWVTOISGRPHTPAGVTVRSNKAETMAVQMLNLEFKLITNTQLSHDKPMGAPNPRG 2775
OY 18 ----- 17
Db 2776 PLKEMAQHEADNDKYAVAAARKYQTOAAMKLSRRMYLIRAMVNGNSEYASNIQOH 2835
OY 18 ----- 17
Db 2836 NVITTSNYPENMLGLVDLYVAQAIHNTGYTSALYITNNACTAVLCGHMDFSCPPDDG 2895
OY 18 ----- 17
Db 2896 WSTSRFLSTHDTITNLADIKSLDSEIAKLDKQEBLSEETAVLKEKRESERIHNOIL 2955
OY 18 ----- 17
Db 2956 RTEDSWTSSEYQVWKGTVIVSVNSGISTESTIEKIMETTGAMFIMCVPITLPSDNTTA 3015
OY 18 ----- 17
Db 3016 HRVGVTSNNVKILTYPGARNTLTPNBEIYVCMKSGRATVQSOTKKDFMISLHFTTVAM 3075
OY 18 ----- 17
Db 3076 IASVQILKVTITFRNEHAMPKALIPSPSSHOSDIEKVVQYPRFTCNFGGSENSLIT 3135
OY 18 ----- 17
Db 3136 TKRRKLVSKALLAKMNERALRHDSNLKDLQAYGRSYLMTQYTERTERYVNIINKNGDELEA 3195

```

```

QY 18 ----- 17
Db 3196 VCIALATMKHRIINBOLBHTMGIMETPRYVDPSITDIVVSCCKNLEKVMSTANTMOEK 3255
QY 18 ----- 17
Db 3256 LGIPASFTGLDELANTQRYVVFIDERTINOSRWMSWSNIRKMSDEGSIKKMASTADNL 3315
QY 18 ----- 17
Db 3316 DMNIGMIOEYRKYLCRKEDIPKTKACPGFKQYVDCAKAQAQOKEARKTDPGARVLVLE 3375
QY 18 ----- 17
Db 3376 EPIDLGIVNDITRAOTGTWTKNENYIRSEITFRASQIRRAQTSRLRSYLDNMNEMVK 3435
QY 18 ----- 17
Db 3436 AYSALBAGDLKVMYRELSRIAIIKSTSTFWHLKEDLEAMDVPNLTRAHELEITKLIG 3495
QY 18 ----- 17
Db 3496 DAVVSRWESKAERKQNEYDFIKSDLGISTLNDHDLNIVLSLAVDRSKVKDLNTL 3555
QY 18 ----- 17
Db 3556 GSQMSVNEAWLRRKIROQCESASNGOLVDHDHANSVSTEQNGKLNILSLHEEA 3615
QY 18 ----- 17
Db 3616 GICDINSDASPLKIRASSLPEAITTCVQTPGMYELTSKVEIITYEAMNDECWRCIE 3675
QY 18 ----- 17
Db 3676 KYVTONIEPHRTINGLRALMOQSKMLTESQAIIVCOLLGNCLLONGETGYVNPAPN 3735
QY 18 ----- 17
Db 3736 KPFVQMLRSODASDYHCYLINLIGADGVRKLSBENIAKENMBOLEHVCVDNKPPIVAVG 3795
QY 18 ----- 17
Db 3796 ENPYACSVHVIDNEDELARLTSQSPYEKIPVLDGREMLPVNNVAMARKTKISSATILK 3855
QY 18 ----- 17
Db 3856 KYPIITGAGSLALTRNYKATPMVTTKPGDVLVLRDCEKKVMCTHTIYDTPGVEQ 3915
QY 18 ----- 17
Db 3916 ILBITDIGGAVCLDGLHCLTPADRTPRDRPVNTSNEVMSLNIETQRYLEATKGLRS 3975
QY 18 ----- 17
Db 3976 HVGQGEKLIHYHNRNEHMYDEHDIYDAMLDENIYIELDPTVQYPTGVNIYNLICIRR 4035
QY 18 ----- 17
Db 4036 MFLAMLXGIPWVAYCDDNPSAQAFLKYYKMGKVFNNNVTTEDLTNTRAMEIINA 4095
QY 18 ----- 17
Db 4096 LFSYQHTTNNHTTRVHLRKAETIILSKEKVSLELNTTIDISVFTAVLGMLENEVILKS 4155
QY 18 ----- 17
Db 4156 QGTGVFDLIPETPEPQLODDMLESKGIIVDIKQCAVYHPTGTVALRVESLSHEINAC 4215
QY 18 ----- 17
Db 4216 STYVKGVDPDEMRAGASNLQSVPASDKADPGSSQLSCSEBMSQGTQPVACVHIIEP 4275
QY 18 ----- 17

Db 4276 CDHANAANDVLEKITEQEQNNHDIISKILATAQAKSNPEDKSMKSGATSADLQATQS 4335
QY 18 ----- 17
Db 4336 LCDSPDGMSQVADQIAKLLEYMNFSEVTLGSIHRORNYVNSQKYLQVDSERMAIT 4395
QY 18 ----- 17
Db 4396 YKGIGEHSGDVPITNPGMMWLAEPFPMKLGIVYKEKEIKVTNNYAGNCAYRLNRKIFREP 4455
QY 18 ----- 17
Db 4456 KSVFTHNVRPEVMDILLYTSANNKQYPTNLQRLADKYLHGRVSHVYVYGEYNDFLOT 4515
QY 18 ----- 17
Db 4516 CYELTRMERVIVTDPADPVYNQOLDADIQVREKIITIKVTSQPEPYRAIMWRFMILFANI 4575
QY 18 ----- 17
Db 4576 DPECELIINGREMAFNDMELVDVWMESEFEDPDIVSALHCOAFNVCCGYAKYHNAKFTD 4635
QY 18 ----- 24
Db 4636 YNTFNDIYEAGNSYSGDEIYMSRYKVKMMWLYIILPGAWYIGIEKMLNVALDTTICLE 4695
QY 25 ----- 24
Db 4696 LKEGETRIDHYKGLFCSANFFHVADPTKGIYDSTVMSVDGIWQYAMOIIINGSHYTE 4755
QY 25 ----- 24
Db 4756 NPVRSHTGSESTGNPNRNISDSEKPMQOLNPNNNMLDPLGLKSEKMIQOFIJDSSVCIK 4815
QY 25 ----- 24
Db 4816 EOSLPIQGLPKPIYVFANQACEAIGVSMNNTGLIYVEVNIQOKLAARNEPQHOTLI 4875
QY 25 ----- 24
Db 4876 LKSHLTLMSTGRHASMKEPYTHLMKADLQNCQYVEGSANSLQMKRETOHAQVYTE 4935
QY 25 ----- 24
Db 4936 ERDVFVRIPQALDHSVGLKLEVKVTGNEIGLSSSGVLARVLKLNRLRANHHLETQH 4995
QY 25 ----- 24
Db 4996 LQKLEFDRRIISHEESSGMTFKPRIGOTLYADVLTAAGIECVNHITTSREWEALK 5055
QY 25 ----- 24
Db 5056 NPLTKVGMPLTNSVLVSGKLGCVPTIPRCYAMLCQGVKLIKFKPETRGGTSNVSAN 5115
QY 25 ----- 24
Db 5116 VGPSNPTTAGPDTPPLITSNMKVYESGASQSPGVSAAKAYQDKLRKKTETPODLRV 5175
QY 25 ----- 29
Db 5176 LETMLFDAYALMSHGLTDLRLNDGNGWHCATATMKSDPEIIEPVGDLPLSINTSAL 5235
QY 30 ----- 29
Db 5236 RHNETNARVIDLMDTDLRDMULTIYAPKNPKLSRVSFGGKINLKTLLTNPCQTRPV 5295
QY 30 ----- 29
Db 5296 PTQVAMGENAVTGRIGSVLPLRREPNNVTHLHKFRTAYYRDGWERVLKDFKANTITIS 5355
QY 30 ----- 29

```

Db 5356 DADVKTMLSRSDMKALATSTIKMETGLSPNPMNAVHVHTESLTKANPIMWTQTOG 5415
 QY 30 ----- 29
 Db 5416 RIIVWQPELCAIAPAIKRIKLELYLDEIYTDGLPDMLSARARTIQDYVEDD 5475
 QY 30 ----- 31
 Db 5476 LVYDQRTDQELIDLEFQVMDLGLDINLMLNRLVHNKRRF 5517
 RESULT 12
 T30226
 polyketide synthase - Streptomyces hygroscopicus
 C:Species: Streptomyces hygroscopicus
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
 C:Accession: J30226
 R:Apuricio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
 Gene 169, 9-16, 1996
 A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
 A:Reference number: Z20782; MUID:96186896; PMID:8635756
 A:Accession: J30226
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-8563 <APA>
 A:Cross-references: EMBL:X86780; NID:9987088; PID:9987100; PIDN:CAA60460.1
 C:Genetics:
 A:Gene: rapA
 C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
 F:54-503/Domain: acetate-CoA ligase homology <ACLI>
 F:1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F:1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 F:3199-3270/Domain: [acyl-carrier-protein] homology <ACP1>
 F:3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F:4787-4858/Domain: acyl carrier protein homology <ACP2>
 F:4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
 F:5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
 F:6760-6831/Domain: acyl carrier protein homology <ACP3>
 F:6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
 F:7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
 F:8412-8483/Domain: acyl carrier protein homology <ACP4>
 F:8482/Binding site: phosphopantetheine (Ser) (covalent) #status predicted
 Query Match 39.7%; Score 62; DB 2; Length 8563;
 Best Local Similarity 0.3%; Pred. No. 1.5e+04;
 Matches 19; Conservative 3; Mismatches 6; Indels 7557; Gaps 6;

Db 780 PDWSTTAASIPVFAFAAYGVLDSLSAGEFVLIHNTAGVGMAATQSAARHLGPRI 839
 QY 11 ----- 10
 Db 840 HATASAKOHILEAGLEDTRIASRTIAFREAFINTDQGVVYNSLSGFVDASLD 899
 QY 11 ----- 10
 Db 900 LLEPRGFLKMTDITRDADRTTADRPCTTQAFDILLADGPPRLRIIAELLEPAQVYL 959
 QY 11 ----- 10
 Db 960 RPLPVLTMIDIRKADAFSWMSRABHTGKITFTIPRLDPDGYVLADGAGALTGVARHL 1019
 QY 11 ----- 10
 Db 1020 VAEQVRRHLLLSRSTPDEALINELIESGARVDTAVCDVSDRAGLVRLAGIAPERPLTA 1079
 QY 11 ----- 10
 Db 1080 VHTGGPAVAHESHQHLRLKGLDLAFAFYFQDAPASVDALARSRAEGLPITAIANGI 1139
 QY 11 ----- 10
 Db 1140 PEAEAVYVRGPLLGRAMASADSAAHIVTRLNTVGLRALAADTLPPLQLVGAHTDTEQ 1199
 QY 11 ----- 10
 Db 1200 QANRQFLAELAEAREQALDLYRVSSVTDLGLSADRYAPDKTSREMGIDSLTVELRNS 1259
 QY 11 ----- 10
 Db 1260 LAATGRLPATLVFPDYPPTPAVLVYRLGELTFGESAPRPARAVSAVQGGEPALIVGMACHL 1319
 QY 11 ----- 10
 Db 1320 PGVSSPEDLMRLVESGTDALISGFPTRGWDVGLFDPDPDASGSKSCYQGGFLDTAAGF 1379
 QY 11 ----- 10
 Db 1380 DASFFGISPREALAMPQQRVLVEVSEAFERAGITPGSVRSDDGVFMGGFPGGAGACA 1439
 QY 11 ----- 10
 Db 1440 DLBGFATAGASVLSGRVSYFFGLBGPATVDTACSSSLVALHQGYALROGECSLALV 1499
 QY 11 ----- 10
 Db 1500 GGVYMTPTQSFVEFSRQGLSADGRCKAYADAADGTEMAEGVLLVERLSDAQAKHQ 1559
 QY 11 ----- 10
 Db 1560 VLAIVGSAVNQDASNGLSAPNGPSQQRVIRALSNAGLAHEDVYVAHGTGTLGDP 1619
 QY 11 ----- 10
 Db 1620 IEAQAVIATYGQDREQPVLLGSLKSNIGHAQAAGSVYIKVMALRHGFVPTLHDEP 1679
 QY 11 ----- 10
 Db 1680 SRHYDSAGAVELVAENRSPATGRPRAGVSFAFVSTNAHYILEGAPQASVDAAAGST 1739
 QY 11 ----- 10
 Db 1740 PYVDSELYPLVSAKSLPALAEVGRRLRAYLAASPGADVRAVGSITLAVTRSLFEHRAVLL 1799
 QY 11 ----- 10
 Db 1800 GDSVTGTGTAVDPRVVEFPGQGMQLGMSALRTSKVFAERMAECAAALSEFVDM 1859
 QY 11 ----- 10

```

Db 1860 LEAVLDDPAVAVDVQPPASMAVNVSLAAVQAACVAPDAVGHSGEIAACVAGAVS 1919
QY 11 ----- 10
Db 1920 LRDAAVVTLSQVYIARGLAGRAMASVALPADVELVDGAWAARNGPASTVYAGAPEA 1979
QY 11 ----- 10
Db 1980 VDRVLAHGAAGVAVRRIADVASHTPHVELIRDELLGVIAGVDSRAPVVPWLSTVDGTW 2039
QY 11 ----- 10
Db 2040 VEGPLDAEYWRNLREPYGEPBAGQLOAGQDVFVEVSASPYLLQAMDDVTVATLRR 2099
QY 11 ----- 10
Db 2100 DDGDAITMLTALAOAYEGVTVDPWPAVLGTTAARVLDLPYAFQHORYWLGVDRAAADG 2159
QY 11 ----- 10
Db 2160 HPLIGTVVALLPGSDGVVLTGRVSLATAMLADHAVRGSVLLPGTAFVELVVRAGEVCD 2219
QY 11 ----- 10
Db 2220 VVDELVIETPLLPGTGQVQMSVSAEADSGHRAVWFSGQADNTDTWTHVTAIVSTSD 2279
QY 11 ----- 10
Db 2280 STVSQPEFAAMPBAPAGAPLDLSDFYDQLTAGYEYGPAGLTAWRDNVTFAEVALA 2339
QY 11 ----- 10
Db 2340 EEOQAQARRAVHPALDALHASNSTLDTAEOGVRLPFSMQVWVHATGPVRLVAIT 2399
QY 11 ----- 10
Db 2400 RTADGMSVLAADSGRPVAVSGLVTRVYATADALGSADELIALTMKEITAPQGTGLTVG 2459
QY 11 ----- 10
Db 2460 RFEDLVSGGVVPEVVVFTALPESSETPLDPLDPLAQTTLTAQVLAQAVQAWLAGBGT 2519
QY 11 ----- 10
Db 2520 NSTLVATGTGLAAGVSGLMRSVQSEHPGFVLYECDDDTLPDOLATVGLDEPLRV 2579
QY 11 ----- 10
Db 2580 CDGRFEVPRILARANTPESSPLTIPDSRAMLLBOPRSGTLBDLALVPTDFAERPLQSGEVR 2639
QY 11 ----- 10
Db 2640 VDVBAAGLNRDVYVVALGMVDRKLAGGENAGVYLEVGEVODLAPGRVFLVGGGFGF 2699
QY 11 ----- 10
Db 2700 SATADRMGLVIPDQMSFTTAASVPVEATAYGYGLVDLAGSGESVLIHAAGVGMMA 2759
QY 11 ----- 10
Db 2760 TQIARHLGARIVASTQKOHILREAGLEDTHTADSRTLSFQETPLNTHGQGVVLS 2819
QY 11 ----- 10
Db 2820 LSGDFVDASIDLPLPGRGRTIEMGKTDIRPHOVTADRPGTTYQAFDLMADGPRLEIIT 2879
QY 11 ----- 10
Db 2880 ELTLFTQVILPLPYQAMDIRQARDAFSMSRAHIGKIIVLIPRRPPDGIILTGGSS 2939
QY 11 ----- 10
Db 2940 GVIAGILARHLAAEHGARHLLLSRTTPDQALIKELAEIGAHVDTATCDVSRAGIARVL 2999

```

```

QY 11 ----- 10
Db 3000 AGVSEPHPLAVIHTAGALDDGVESLTQQDVTILPRADGAWMLHLLTONTDLAAFWM 3059
QY 11 ----- 10
Db 3060 YSSAGVLSAGQGNVAAANAFVDALAEQRGRCULPALAVANGIMEDTSGLTAKMTDTR 3119
QY 11 ----- 10
Db 3120 DRIIRGLRAISAGRGMLDAAASHRGEFVLLAASMEPVDEVPALLRLHRPVARRAA 3179
QY 11 ----- 10
Db 3180 STGDSSVOMLARIAPYEREKALKLYCGAATVYLGHADASTIPATGAFEDLGVDSTAVE 3239
QY 11 ----- 10
Db 3240 LRNGLAKATGLRLPATLVFDYPTPALAARLEBELFTGENPAPVRTSVSVVQDEPLATVG 3299
QY 11 ----- 10
Db 3300 MACRLPGVSSPEDLMRLLESCTDAVSGPTDRGMDVENTLPGPAGNSYRLQCGFLDAAA 3359
QY 11 ----- 10
Db 3360 GPDASFPGISPREALAMPQQRVLVEVSWEAFERAGIKFVSRTGTGTGVFKAGYPGGYGI 3419
QY 11 ----- 10
Db 3420 GADLGGFTTAGAVSVLSGRVSFFGEGPAPFTVDTACSSSLVALHQAYALROGECSLA 3479
QY 11 ----- 10
Db 3480 LVGCVTVMATPQTFVEFPAQGLADGSKAPADSADAGFSEGVLLVERLSDARRNG 3539
QY 11 ----- 10
Db 3540 HQVLAVVRGSAVNQDGSNGLTPNGPSQQRVIRAAALSNAGLSTAEVDVVEAHGTGTLG 3599
QY 11 ----- 10
Db 3600 DPEAQMAMATYGODREOPLILGSVKNLGHQAAGVGIKVMALORGFVPTLHVD 3659
QY 11 ----- 10
Db 3660 EPSRHVDSAGAVOLVTENQPPGTDRPRRAGVSSFGISGTNAHYILESALPTOPAGNTV 3719
QY 11 ----- 10
Db 3720 VESAFEWVPLVISTQSALEVEGRRLRAYLAASPGADTRAVASTIANTRSVEYRAVLI 3779
QY 11 ----- 10
Db 3780 GDDVTGTAAANDPRVVFVPGGSGQAGMBELAAFPVFAIRHQVMDLIDVDLVNE 3839
QY 11 ----- 10
Db 3840 TGYAPALFALQVALFGLLESNGVAPDAVGHSGVELAGTVSGIMSLDCACTLVASARAR 3899
QY 11 ----- 10
Db 3900 LMQALPAGVMAAVPSEDEARAVLGEVEIAAVNGPSSVYLSGDEAAVLQAAAGLGKWT 3959
QY 11 ----- 15
Db 3960 RLPTSHAFHSAEMPELFEFPAVABGLTYRPPQVAMAGDQVMTAEYVWQVDRDTVFG 4019
QY 16 ----- 15
Db 4020 QVASFEDAVFVELGADRSLARLVDSIAMLHGDBHQAAGALAHLYVNGVSEMSAVLGD 4079

```

QY 16 ----- 15
Db 4080 VPTRVLDLPTTAFQHORWLEGTDRATAGCHPLLSVRLAASGVLTARVSRGDL 4139
QY 16 ----- 15
Db 4140 LRQDTVLPAVFEVMALAADEVCGGLVEDLSVEALLLPDDGAVEQVQWGPDEGR 4199
QY 16 ----- 15
Db 4200 RLVSFARYGDEPWTCLATLATTGVAAGVAGVMPAGAVPGTSAPSLRAV 4259
QY 16 ----- 15
Db 4260 RLGSDFAEVALDHAHATRFVLPALMAAALLTVEETPAVMQGLLHAGNGELRYL 4319
QY 16 ----- 15
Db 4320 TSHDGLSAEATDSTGLPYLLARSLLRTVPYEPATSDLLTLWAGIPTPOQTGLT 4379
QY 16 ----- 15
Db 4380 VGAFEDLAADGDVPEVAVFTALPDSDPLEQTRKLLTAQVLTQLQEMLGERSSTLY 4439
QY 16 ----- 15
Db 4440 VRTGTGLAAAGVSGLMSAOSHEHGRFVIVESDALLQODLAAVGLDEPRLRYSGRYE 4499
QY 16 ----- 15
Db 4500 VPRLTRHAEPERPTWDPDQVLLITGSGVLGIAARHLVTERGVHLLLSRSAPDE 4559
QY 16 ----- 15
Db 4560 ALIGELGARVETACDVSDPALTOVLACVSEPHLTAVITAGVNDGVVESTLYQ 4619
QY 16 ----- 15
Db 4620 RLETVLRKADGAMNHELTRDADLAFAVMTSSAAGVLSGAGQANVAANAFDLALAEOR 4679
QY 16 ----- 15
Db 4680 HAEGLPALAVWGLMEDASGLTAQLTDTDRIRRGILRAISAEHGMGLDSASRHEPV 4739
QY 16 ----- 15
Db 4740 LVAAPEVRDAEVPALLRSILHRPIARAAAGARWLAALAPAREKALLKLVCSAAT 4799
QY 16 ----- 15
Db 4800 VLGHADTSTVSAAVFRDLGVDLSLAVELRNSLAKATGLRLPATLVEDYPTPALAVRLG 4859
QY 16 ----- 17
Db 4860 ELFTGENPVKRPVSAVAODEPLATVGMACRLPGVSSPEDLWRLLESTDAVSGFPTD 4919
QY 18 ----- 17
Db 4920 RCMDEVNLYDMAGSKHRAEGFTLDAAGFDAGFPGISPREALAMPQORLVLEWSMEAFE 4979
QY 18 ----- 17
Db 4980 RAGIEPGRVSGSDTGVFMGAVPGYGAGADLGFAATASATSVLSGRVSYFEGLEGPAFT 5039
QY 18 ----- 17
Db 5040 VDTACSSSLVALHQAQYALRQEGSLALVGVTWATPELFTERSKORGLASDGRKAF 5099
QY 18 ----- 17
Db 5100 DSADGTMAEGVGLVLRSLSDAQAKHQVLAVVRSSAVNDGASNGLTAPNGP9QORVI 5159
QY 18 ----- 17

Db 5160 QAALSNAGLAHEVDVVEAHGTGTLGDPTEAQAIVATYGGDRERPLLGLSKNIGHAQ 5219
QY 18 ----- 17
Db 5220 AASGVSVIKVMALQHMTVPRTLHVDEPSRHHVMAGAVELVRENQWPGTDRPRRAGV 5279
QY 18 ----- 17
Db 5280 SSFGVSGTNAHYILSAPPAQPAEPAQVETPVASDVLPVLSAKTQPALTEHEDRLRA 5339
QY 18 ----- 17
Db 5340 YLAASPADTRAFASTAVTRSFEBRAVLGGDTVTGTAVSDPRVVEVFPQGNQWLMG 5399
QY 18 ----- 17
Db 5400 GSALRDSVYFAERMAECALSEFVMDLVLDPPAVDVRVYQPA5MAVMVSLAAVW 5459
QY 18 ----- 17
Db 5460 QAAGVRPAVIGHSGEITAAACVAGVSLRDAARIVTLRSQA IARGLAGRMASVALPA 5519
QY 18 ----- 17
Db 5520 QDVELDGMIAHNGPASTVIAGTPEAVDHYLTAEARGVRRTIVDASHPPHVELI 5579
QY 18 ----- 17
Db 5580 RDELDDITSOSSQAPLVFWLSTVDG5WVDSPLDGEYTRNLREPVGHFPAVGQLQAQGD 5639
QY 18 ----- 17
Db 5640 TYFVEVASPVLQAAMDVVTVATLRDDGDATRLTALAQAYHGVTVDPALLIGTTT 5699
QY 18 ----- 17
Db 5700 TRVLDLPTTAFQHORWVEGVDSAGGHPLLGVAVELPD5NGVVLTRGVSLATHTWLAD 5759
QY 18 ----- 17
Db 5760 HAVRGSVLLPGTAFVLLVRAADVECDVYDELVIETPLLEPQTGVQLSVSAEAD5EG 5819
QY 18 ----- 17
Db 5820 HRAVTVSRADNADTWTNRHSATISASDAPLSLPEFASWPPAQOPTNYGDLYDRLAAG 5879
QY 18 ----- 19
Db 5880 TEGPAFQGLQAAMWDDGTAYAEVALAEQOAEARFAVHPALLDAALHASVLTHTPDAEQ 5939
QY 20 ----- 23
Db 5940 QSLRPFMSHQVYHATGSATLKVAMTPTTDG5VHADDGRPVATIGSFYTRPVYADA 5999
QY 20 ----- 23
Db 6000 LGSAADLLRVWTEIPQQTGLTGRFEDLVADADVPEVWVYVYARBDTGSDDLQAQ 6059
QY 24 ----- 23
Db 6060 TRTLTAQVLOAVQAMLAGERFTDSTLVRTGTGLAAAVSGLMRSVQSHPRGFVLVSGD 6119
QY 24 ----- 23
Db 6120 DDTLTPDQLAATAGDEPRLRVCGRFVEVRLARANTPESSPLTTPDDRAMLLEQPRSGT 6179
QY 24 ----- 23
Db 6180 LODIALVPTDARPLRPGEVRLDVRAAGLNFBDVLIALGTYPGEAVIGAEAGVVL5EVG 6239
QY 24 ----- 23

```

Db 6240 PEAHDLAEGDRVFGVGGFGAVAIADRMLAVIPDGMSFTTAASVPVFAFANYGLVDL 6299
QY 24 ----- 23
Db 6300 GGLSAGESVLIHAAGVGMAQTARHLGAQYATASAGKHILYEAGLDGTRIADRT 6359
QY 24 ----- 23
Db 6360 TGFREAFINTDGRGVVNLNSLGDVYDASIDLPRGGRVEIGKTDIRDPHRITADP 6419
QY 24 ----- 23
Db 6420 GTTYACDLMDVGPRLKEITTEILSLFGQVLPPLVQTDIRQARDAFSGMSRAHIG 6479
QY 24 ----- 23
Db 6480 KIVLTIPRRPDDGITLITGSGVLGITARHLAEHGRHLLLSRTAPDEALIKELAE 6539
QY 24 ----- 23
Db 6540 LGARVETACDVSDRAGIARVLAGVSPHPLTAVHTAGALDGVESLTTQQLDTVLRP 6599
QY 24 ----- 23
Db 6600 KADGAMHLELTRADLAFAVYSSAAAVLGNEGQNTAAANATLDALAEORRTQGLPAL 6659
QY 24 ----- 23
Db 6660 ALANGPWEYTDLTQAOLGTGDRIRCSGMRTTAEDGMRLFDASHGEPILVPAVLP 6719
QY 24 ----- 23
Db 6720 TRDGEVPALESLRPIARRAASADGVQWLALAPAREKALKVCSAAVYLGHADA 6779
QY 24 ----- 23
Db 6780 RSIPAGAFKDLGVDLSMAVELRNGIVKATGLRLPATLVVDYPTPIYLARLDLPTGEN 6839
QY 24 ----- 23
Db 6840 PAPVRGPVSVGDEPLAIVGMACRLPGVSSPEDLRLVESGTDAVSGFPTDGMDEVN 6899
QY 24 ----- 23
Db 6900 LYDSDPEAGSYCVGGFLDTAGFDAGFFGISPREALMDPQGRLLLEVSWAEFRAG 6959
QY 24 ----- 23
Db 6960 IEPGVSVDYGVFICAFPVYCGAGFDREGYATSGPSVLSGRVSYVGLGPAITMDTA 7019
QY 24 ----- 23
Db 7020 CSSSLVALHMAQALRNGECSMALAGSVTVAPEVTEFARQGLASDRCRAFDASD 7079
QY 24 ----- 23
Db 7080 GAGSEGAGLLVERLSDARRNHQVLAAYRGSANVDGASNGLTAPNGSPQORVIRAL 7139
QY 24 ----- 23
Db 7140 SNAGLSTADVVEAHGTGTLIDPIEAQALLATYGGDRBQPLLSKSNIGHTQAASG 7199
QY 24 ----- 23
Db 7200 VSGVIRKVMALRIGFVPRITLHDEPSRHVDMAAGAVELRENOQWPGTDRPRAGVSSFG 7259
QY 24 ----- 23
Db 7260 VSGTNAHVLESAPPAQPAEEQPVETPVVASDVLPLVISAKTQPALTEHEDLRATYLA 7319
QY 24 ----- 23
Db 7320 SPGADTRAVASTLAATVRSVEHRAVLLGDGDAVTGTAVTDPVVVFPGQGWQWLGMSAL 7379

```

```

QY 24 ----- 23
Db 7380 RDSVFAERMAECALASEFVMDLFAVLDDPAVYDRVDVQPASMAVWSLAAWQAA 7439
QY 24 GIRP----- 27
Db 7440 GVRPDAVIGHSGEIAACVAGASLBDARITYLSQATARGLAGRAMASVALPAHEI 7499
QY 28 ----- 27
Db 7500 ELVDGAMIAAHNGPASTVAGTPEAVDHYLTAHANGVRVRIITYDYASHTPVELIRDE 7559
QY 28 ----- 27
Db 7560 LLGITAGISQPPVPMVLSITVDSWVDSPLDGEYWRNLREPYGFHPAVSQLOAGDAVE 7619
QY 28 ----- 27
Db 7620 VEVSASPVLLQAMDDVYVATILRRDGDATRLTLAQAIVHGTVDPAILGTTTARV 7679
QY 28 ----- 27
Db 7680 LDLPYAFQHORVYKSVDRAADGPHLLGAVVELPESDGLLTGRVSLATHAWLADHAV 7739
QY 28 ----- 27
Db 7740 WGRVLLPGTAFVELVYHAAGEVCCDVYDELVTETPLLPTQSGVOLSVSGEADSGHRV 7799
QY 28 ----- 27
Db 7800 VTFESRADNADTWRHVSATVRSVDTTPPSDLTAMPQAQKPVYAGFYDOLTGMYEY 7859
QY 28 ----- 27
Db 7860 GPAFGQLQAAWRDGTVEAEVALAEQVREARAYAVHPALDLAALHACTLNASDAEYGVG 7919
QY 28 ----- 27
Db 7920 LPFSWGVVHAGSAMLKVAVTQAADGWSVYVADDIGRPVASVSLTVPYTDALGSA 7979
QY 28 ----- 27
Db 7980 ADDLLATWAGIPTPQOTGLTVGRF 8004
QY 28 ----- 27
RESULT 13
T20802
Hypothetical protein F12P6.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Kershaw, J.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19326
A:Accession: T20802
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1059 <WTL>
A:Cross-references: EMBL:Z73425; PTDN:CAA97790.1; GSPDB:GN00022; CESP:F12P6.5
A:Experimental source: clone F12F6
A:Map position: 4
A:Introns: 57/1; 437/3; 741/2; 984/2; 1022/2
Query Match
Best Local Similarity 39.1%; Score 61; DB 2; Length 1059;
Matches 14; Conservative 8; Mismatches 8; Indels 699; Gaps 4;
QY 1 SRX----- 4
Db 234 SRHRSLVKYFRKREKEFVVRKCTKARNEYLLCVKANAAALHFFAQLSLIDCMDL 293

```

```

Oy 5 -----XHSME----- 9
Db 294 GMDFWLLEKYEYDERKKTTOHEMDSLASLSTRSSVDYKADKOFPEANHOLEMPLPKO 353
Oy 10 ----- 9
Db 354 FERRPOLGDDIMEVSAQSSLDLQRLQLEKRLBGLQFEVDEWKSLEASEKOLLQLY 413
Oy 10 ----- 9
Db 414 NTFDGEAGKWRNDLHTVYQYLLKKEFFLLNGMLERLEARSASIGELAKHGSLKSA 473
Oy 10 -----XRTDP----- 14
Db 474 SVGSASERNSTFTENGSGONHLLERRSKARIGCIYTKSPDDRPRPKFGSLDEYV 533
Oy 15 ----- 14
Db 534 EATGEIPLVQSAIAYLSRYSLNGLFRVSGSQSEINFRAYERGEDLLFQYLDGSD 593
Oy 15 ----- 14
Db 594 ANSAGVILKLYFRELRPEIPFIFMEQPCDCANSESPTEFVRARRELVSRLPVSHVILLR 633
Oy 15 ----- 14
Db 654 FLFAFLSHLCEPADENMEPHNLALICFPPLLPFPEGKQVYFHYNYVELVRLNLIHAD 713
Oy 15 ----- 14
Db 714 VFPRDLGAVYDYKAMORYMDGNFIEENDLISEDEAHEKLSLPSRHMMIDSTYESADR 773
Oy 15 ----- 14
Db 774 ILLSPILSQANTSTPNGISPSNGAESSTHNEIDYAPLASRSSNRKASGIDLAMRSE 833
Oy 15 -----INPAMYXK----- 23
Db 834 MPHRIANELNFIKNSLSSESGKISVLRSHVPEPMTDNRHLNMRSMSTDPDEREY 893
Oy 24 ----- 23
Db 894 VSPPEPLSTKYVSAASCISAPARTIDIKERKDRDLXAPITRATVPANSGSNTEN 953
Oy 24 --GIRPVGR 30
|:|:|:|
Db 954 STGVAPICK 962

```

RESULT 14

```

B96695
hypothetical protein F5A8.4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence: revision 02-Mar-2001 #text: change 31-Mar-2001
C:Accession: B96695
R:Thellogis: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.M.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B96695
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-5138 <STO>
A:Cross-references: GB:AE005173; NID:94204276; PID:AAID10657.1; GSPDB:GN00141
C:Genetics:

```

```

A:Gene: F5A8.4
A:Map position: 1
Query Match 39.1%; Score 61; DB 2: Length 5138;
Best Local Similarity 0.7%; Pred. No. 7.5e+03;
Matches 17; Conservative 4; Mismatches 10; Indels 2303; Gaps 4;
Oy 1 SRXHX----- 5
|:|
Db 1439 NLLHTRGLTVRDLLSWVAFVMMATESIGPAVALIHGAFLVLDGLSLGTGSGRDGDL 1498
Oy 6 ----- 5
Db 1499 REKCEAFLLQLELEFASDPLPLELSRMELYGWDSKAIKESKSVRHEGMGIDPFISK 1558
Oy 6 ----- 5
Db 1559 GDENPEIGFEFLAFTTHRVNLRLVLRAMOLSKPILBSPGVGKTSLLALGKYSGHKV 1618
Oy 6 ----- 5
Db 1619 RINLSQTDMDLGLSGLPVESEDMKFAWSDGILLGLNALIDHRAQVFIPELGCTPEC 1678
Oy 6 ----- 5
Db 1679 PTFRVFAQNPSTGGGKGLPKSFLNFKVYVDELVEDDYLFICSLYSPVSPILS 1738
Oy 6 -----HSM----- 9
|:|
Db 1739 KLIALNROLHDGTMLYRKGHGSPWFENLRVIRSCQFQEAHIDLEVSEFLVLTQR 1798
Oy 10 -----XRTDINP----- 17
|:|
Db 1799 MRTATRKKEVRLIYKAIFDKTSPINDYPRVQLPAPVAVGTAAIKRLNLSNASEQLK 1858
Oy 18 ----- 17
Db 1859 LPEIKONLEVAHVQONKMLCILVGPSSGKTSYRIILAQLTGYPLENLNSATDSDL 1918
Oy 18 ----- 17
Db 1919 LGCFEQYNAFRNRLVMTREVEHLVDEYNLSLLQSSQELFNSRGLVSRMLSTYLNKIDS 1978
Oy 18 ----- 17
Db 1979 LVENPLFLINDSETLSTLEEVEDLEQVLKEGVLVPSMKTLYLEQISKTIIQLQTHEKQ 2038
Oy 18 ----- 17
Db 2039 STKEFWTGMKLIKAEKGEVYVLLKANLCPVYLDRLINSVPEGSIITNEGIVNGEPV 2098
Oy 18 ----- 17
Db 2099 TVVPHNPLFLSVNPKFGEVSRAMNRNGVEYFMGPHQWLNEDGSNCEBELVGRVERPL 2158
Oy 18 ----- 17
Db 2159 ALSIGPKYIVTSMAAAHVAMVINGSGFVRIYLELEQWVHLFOLLNNGNQLIMSLQL 2218
Oy 18 ----- 17
Db 2219 SMEHYILSLGVTGKEVYDFVRETYLSDELSELDSFNGDLYLPGGPKRFNLRDLTW 2278
Oy 18 ----- 17
Db 2279 YSRRTYRONCMYLEFLGAQYASHQPKISDNVKSRLRELAAGEPRITYSIDSWTKVLF 2338
Oy 18 ----- 17
Db 2339 PKALIGSSCAPDAANFENDLASKMLLFAANWTIQATEEDIQIYLAWSFGSRLQOHC 2398
Oy 18 ----- 17

```


Db	2399	FLIJCFLNTLKV	EEHPIINNHSIRCKNKLKLCRLDPAV	IPMLSSKLIDVAASNDQSR	2458
QY	18	-----	-----	-----	17
Db	2459	YKSLFESLNSVGL	LRSTYOQMLVESNDNHTDSTFRFLDS	RLVLEKKILCEIVGADSF	2518
QY	18	-----	-----	-----	17
Db	2519	SVLIQLYTEVIDNHS	FFWMSGVSSDDEYLLFSPWSL	IKIKMHSPFGEVOVLEESKN	2578
QY	18	-----	-----	-----	17
Db	2579	INNVLHGHPEKSM	LWAYGHPSLPVSABELFKQOEFLQ	LCSTWVPLKSEDEHGNHLT	2638
QY	18	-----	-----	-----	17
Db	2639	KAIPESGECLLAL	BEGLCISSTIADDDVDYAAVOLD	EYIQTLERLKLEKKRLEDKM	2698
QY	18	-----	-----	-----	17
Db	2699	GFSEIDENTENTASC	VFCEPEIYVTGSGFSWKTCTFAS	SECSGLDYELLAAQHLLVA	2758
QY	18	-----	-----	-----	17
Db	2759	RPTEHODLVDIRKL	KPALLEYSLSTRPOTLVAHQKLMA	IDAHASELGVDTKIAGPAL	2818
QY	18	-----	-----	-----	17
Db	2819	EIYTWHSVLAKN	QSIGLMAHLVTSFPHLMFTGPT	EYLRHMQLSDSVTFADSNCENSY	2878
QY	18	-----	-----	-----	17
Db	2879	RCSDSIYTHOKSE	SETFEVAIKSFVHALEKQNMKD	GIOMLISLIGSSHNKLSTVHS	2938
QY	18	-----	-----	-----	26
Db	2939	FVGPLAKRLYS	DSSSNEFYCNGLAMLYIGGLR	PHLNSLDVIDPAMTKLKLKEKI	2998
QY	27	-----	-----	-----	26
Db	2999	SSELEINIVRGEC	GYISGLLYSGNDESEHTLSKLT	EKHRLQKVIFRSDPRKYQDLR	3058
QY	27	-----	-----	-----	26
Db	3059	RALDEFAGFLR	PLRPSLVNDIEVLDMQVVEQVFN	NOETALISFIDRMSSDYSEYDITOPT	3118
QY	27	-----	-----	-----	26
Db	3119	QVSYTEMKLGL	SLFVSGALLGKLINFRDIDM	DSVETIYALMRPRDSSIASFTYTECL	3178
QY	27	-----	-----	-----	26
Db	3179	PPLHLSHGANS	RKAKSLGLDVGLLHLKLI	SVSAEDSRKASEQLVALYKNLHARVLOFVA	3238
QY	27	-----	-----	-----	26
Db	3239	NTGILDEAS	FELDKIYELARIMEMKQAKT	KADNLPGLYKFRSDPKIDSMEYDIS	3298
QY	27	-----	-----	-----	26
Db	3299	ALGTYFPNESF	SEMEOYLADDDTKNWKMTHTID	DEENLEDDMDLIQBLHDSIYSTHNL	3358
QY	27	-----	-----	-----	26
Db	3359	FGFCDLSEKTI	YISIQSGRCITDSRRLDSFT	DSYBELGYSMTKGLRGLFTSSDAKLVPEH	3418
QY	27	-----	-----	-----	26
Db	3419	ILRLCLENKKN	FTSNYQASAKNYFYKDLDB	PELGMVFTPLQORINSILOEREDHPL	3478
QY	27	-----	-----	-----	26
Db	3479	QKLSGVLOMLAI	PSSTPLAKALSGLOFLCKV	AKLOEGCKPLISDLEPITISLASSMO	3538

```

QY 27 ----- 26
Db 3539 KVEFERMPLLDEVDQOYELNARKFCFRMLMKFQNMKTSPPHKVSFLDKLGLOHYILK 3598
QY 27 ----- 26
Db 3599 SSQLIAPLELQLTSLNLYFVDSRFYDSDSHKRRVEMCNINFGYILOPLVPMEOQLDILNR 3658
QY 27 ----- 26
Db 3659 KNEVELKEVCLKICRWERPDNYLYNETYKTRRQKVKLIKQFTDMLRLPVMVKPDLTKE 3718
QY 27 ----- 31
Db 3719 RAOFLPLDDPLMDGASDMRIVLVASLDAEDQLDRSSWYVVMNKLKESVGRF 3772

RESULT 15
A88852
protein unc-22 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: A88852
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:title: Genome sequence of the nematode C. elegans: a platform for investigating bid
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.
A:Accession: A88852
A:Status:"preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <SNO>
A:Cross-references: GB:chr_IV; PIDN:CAA98081.1; PID:g3681830; GSPDB:GN00022
C:Genetics:
A:Gene: unc-22
A:Map position: 4
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology

Query Match 39.1%, Score 61; DB 2; Length 6831;
Best Local Similarity 0.3%; Pred. No. 1.3e+04;
Matches 16; Conservative 6; Mismatches 9; Indels 4954; Gaps 4;

QY 1 SRXHSME----- 9
|||:::
Db 1127 SRNGHTILKIKIEMNEADYIEIDQAGLRGSCNTVLEAEKRPILNMKPKRIEAKAGEPC 1186
QY 10 ----- 9
Db 1187 VVKVPEJIKGTRRQDPKAKQILKNKRPIDEMRKLEVEIINKDVAEIVFKNPQLADTGKWA 1246
QY 10 ----- 9
Db 1247 LELGNSAGTALAPPELFVKDKPKPKGPLETKNWTABGLDLMGTDPPEDEGAPVKAYIIE 1306
QY 10 ----- 9
Db 1307 MOEGSGMMAVGETKGTDFVKDKLKEHEGYKFRYKALNECGSLDPLTGESVIAKNPYGV 1366
QY 10 ----- 9
Db 1427 TDKRVVEDKEYLVNKAANKAGPDGDHCKPIMKAKKASPEFTGGIKDLRLKVGETI 1486
QY 10 ----- 9
Db 1487 KYDVPISGEPLPECLMVMYNGKPLKAYAGRYKMSERGHIMKIENAVRADSGKFTITLKN 1546
QY 10 ----- 9

```

Db 1347 SGSCDSTATYVVRPTPKPGLIDADVCAATLSNNPPDDGGDPLTGYIYEADMDN 1606
QY 10 ----- 9
Db 1607 KGYIEVGKVDPMNTTLKVNGLRNKNGYKFRVAVANNEGESEPLSADQYQIKDPWDEPG 1666
QY 10 -----XRPDIN----- 16
Db 1667 KPGREITDADRIDIAWEPHKGAGPIEYIYEVRDPTKEMKEVKVPTDINASISG 1726
QY 17 -----PAW----- 19
Db 1727 LKEGKEQFRVAVANNAKGPQSPSEKQIAKPKFIPAMLKHDNLKSITVAGATVMEV 1786
QY 20 ----- 19
Db 1787 KIGEPPIPEVAKFKGNQOLENGIOITIDTRKNEHTILCIPSAMSDVGEYRLTVKNSHA 1846
QY 20 ----- 19
Db 1847 DEKANLTVLDRPSKPNGLFVSDVEEDNLSMKPPDDGGEPIEYVEKLDATGRW 1906
QY 20 ----- 19
Db 1907 VPCAVKDTKAHIDGLKKGQTYOFRVAVANKEGASDALSTDKTKANPYDEPGKTGTPD 1966
QY 20 ----- 19
Db 1967 VVMDADRSLMEBPKSDGAPITQYIEKKHGRDQMGCVSGDQNAETILGLEK 2026
QY 20 ----- 19
Db 2027 EYQFRVAVANNAKGEASDSRKVAVANLKPWIDEAMKTTIKVGNVEFDVYRG 2086
QY 20 ----- 19
Db 2087 EPPKEMIFNEKEPYVDOKIRIESDYKTRFVLNGATRKHAGLYTLATNAGSDKHSVE 2146
QY 20 ----- 19
Db 2147 VIVLCKPSPGLPLEVSNVYEDRADLEWKPEDDGAVIDHYEIEKMDLATGRWPCGRS 2206
QY 20 ----- 19
Db 2207 ETTKTTVNLQGHYKFRVAVANKEGSDPLTTNTALIAKNPYEVPKVDKPELVMDK 2266
QY 20 ----- 19
Db 2267 DHVLAWNAVDDGAPIEAFVIEKKDKNGRMEALVPGDKTATVPLKEGEYQFRIS 2326
QY 20 ----- 19
Db 2327 ANKAGTDPSPDRVAVANLAPRIHREDLSDTYKVGATLKFIVHIGEPAPDVTW 2386
QY 20 ----- 19
Db 2387 SFNGKIGESKAQIENEPYISFALPKARKSGKTTITATNINQSDSVTINIKVSKPT 2446
QY 20 ----- 19
Db 2447 KKGPIEVTDFEDRATLDMKPEDDGGEPIEFYIEKMKNTKQGIWPCGSGDTHFTVD 2506
QY 20 ----- 19
Db 2507 SLKNGDHYKFRVAVANSESPDLETFNDILAKNPFDRDRGRREPTDMSDHDLDKMD 2566
QY 20 ----- 19
Db 2567 PPLSDGAPIEYQIEKRTKYGWEPALIVPGQTTATVPDLTPNEEYEFVAVANNGKP 2626
QY 20 ----- 19

Db 2627 SDPSASAKAVIAKPRNLKPHIDRALKNLTIKAGOSIFDVPVSGEPAPVTWHPDNR 2686
QY 20 ----- 19
Db 2687 IRNGRVKLDNEPYOSKLVYKOMERGDSGTITIKAVNANGDEATVKINVIDKPTSPNGP 2746
QY 20 ----- 19
Db 2747 LDVSDVHGHTVLTNWRAPDDDGIPLENYVIEKYDTASGRWVPAKAVAGDKTAVVDGLI 2806
QY 20 ----- 19
Db 2807 PGHEYKFRVAVANAGSESDPLETFGTTLAKDPDKPGTNADETDMDKDHVLEMKPPA 2866
QY 20 ----- 19
Db 2867 NQGGAPIEYVEMKDESPFNDAVHAPAGOTNATVGNLKEGSKYEPRIRAKNAGLGD 2926
QY 20 ----- 19
Db 2927 PSDSASAVAKARVVPVIDRNSIOEIKVAGODFSINIPVSGEPTTITWTGTPVESD 2986
QY 20 ----- 19
Db 2987 DRMKLNEDGKTFHVKKALRSDDTGYIILKAEENGDTAEYKTVYLDHPSSRGPDLVT 3046
QY 20 ----- 19
Db 3047 NIVKDGCDLAWKPEDDGGAETSHYIEKQDATGRTAGSGSKDTNFHVDLTOGHEYK 3106
QY 20 ----- 19
Db 3107 FRVAVANRHGSDPLEAREALIAKDPDRADKPGPEIYDMDKHAHLKMTPRADDGAP 3166
QY 20 ----- 19
Db 3167 IEGYVEMKTPSGDWVPATVVGAGELTATVDGLKRGQTYOFRVAVANKEGSDPDSRT 3226
QY 20 ----- 19
Db 3227 MNAKPHLAPKINRDMFAORVAKAGOTLNFVDVVEGEPARKIEMFLNGSPLSSGNTIID 3286
QY 20 ----- 19
Db 3287 NNTDNTKLTSTABASGKYIVATNKGDEHVDVNIIDIPAGPGLRHDITKE 3346
QY 20 ----- 19
Db 3347 SVYIAKDEPLDDGSPITNTVVEKQEDGGRWVPCGETSDTSLKVNKLSGHEYKFRVAV 3406
QY 20 ----- 19
Db 3407 NROGTSAPLTSDBAIVAKNPFDEPDAPTDVTPVDMDKDHVLEWKPANDGAPIDATYV 3466
QY 20 ----- 19
Db 3467 EKKDKGDMWECARVDDKTTKATADNLTPGEITYOFRVAVANNAKGPCKPSDPTGNVYAKR 3526
QY 20 ----- 19
Db 3527 RMAKLNLAGLIDLRKAGTPIKIDIAFEGEPAPVAKMKANDATIDTGARADVTNTPTSS 3586
QY 20 ----- 19
Db 3587 AIHIFSAVRGDTGVYKLIIVENHGDQAQCNVTVLDPGTREGPLKIDELHKEGCTLNNK 3646
QY 20 ----- 19
Db 3647 PPTDNGTDLVHYIYEMKDTSRGTWQEVGTFPDCIAKVNKLVPGEKAYRVRVANLQGES 3706
QY 20 ----- 19
Db 3707 KPLAEERPIIAKNOFDVDPVDKPEVTDMDKDRIDIKMNPANNGAPVYGYIYEKKEKG 3766

QY 20 ----- 19
Db 3767 SAITBAGKIPGTTESADNLKPGVEEYEVYAVNAGSPDSDPTDQITKARYLKPKIL 3826
QY 20 ----- 19
Db 3827 TASRKIKKAGFTINLEVDFIGADPTATMTVSGAALAPELLVDAKSSSTTSIFFPSAK 3886
QY 20 ----- 19
Db 3887 RADSGNKLKYNMELGDEAIFEVYQDRPSAEGPLEVSDVTKDCSVLNMKPKKDDGA 3946
QY 20 ----- 19
Db 3947 EISNYYEKRDKTNTWVPVSAFVGTSTIVPKLTGHEXEFRVMAENTFGSDSLNTDE 4006
QY 20 ----- 19
Db 4007 PVLAKDFGTGPKRGREIYDNDHIDIKMPPRONGSPVDHYDIERKDAKTGRITKY 4066
QY 20 ----- 19
Db 4067 NTSVQGTASTDRVOKGHTYEVYAVNAGSPDSSAATAKPMHEAPKFDLDDG 4126
QY 20 ----- 19
Db 4127 KEPRVAGEPLVITIPFTASPDIDISWTKEGKPLAGVETTSQTKLIVPSTRSDSGPV 4186
QY 20 ----- 19
Db 4187 KIRAVNYGEAEANIKITVIDKPGAPENITYPAVSHTCTLMNDAPKDDGACIAGYKIE 4246
QY 20 ----- 25
Db 4247 YQEVGQIMDKVPLISGATYVGLGHEGQYFRIRAEANAGLSDCGVVYIKDPPD 4306
QY 26 ----- 25
Db 4307 PCGAPSTPEITGYDINOVSILAMNPPRDDGSPILGIVVEREFERKGGDMAVYKMPVKGCT 4366
QY 26 ----- 25
Db 4367 ECIVPGLHENETYQFRVAVNAGHGEPSNGSEPTCRPYVEKGPADAPRVGKITKNSA 4426
QY 26 ----- 25
Db 4427 ELTWNRPRLDGCAPIDIGYVEKKKLGDNDWTRCNDKPVBDTAEVYNLGEKEEYEFVYA 4486
QY 26 ----- 25
Db 4487 VNSAGGEPSPKPSDLVLIBEQGPRIFFDINNLKIDTVRAGETIQIRIYAGGNPKPIIDL 4546
QY 26 ----- 25
Db 4547 FNGNSPIFENERIVDVNNGEIVITTTGSKRSDAGPYKISANWKYKDTCKINVEVLADP 4606
QY 26 ----- 25
Db 4607 GKPTGIRATDIOADAMTISMRPKDNGDALTNNYVERKTPGDMVTVGHPVGTTLRVR 4666
QY 26 ----- 25
Db 4667 NLDANTPEFRVRAENQYGVGEPLFETDIAIVAKNPFDPGAPGQPEAVETSEALITLQMT 4726
QY 26 ----- 25
Db 4727 RPTSDGAPIQGYVIEKREVSTEWTKAAGNILDTHKRVGLTJPKTYEFRVAVNAG 4786
QY 26 ----- 25
Db 4787 QGEYSVNSVPTADNAPTRPKINMGMLTRDILAYAGERAKIIVPAASPAKVTTSKGEN 4846

QY 26 ----- 25
Db 4847 KISPTDPRKVEYSDLELATLIEKSELTDGGLYFELENSQSDSASIRLKYVDKPSAQ 4906
QY 26 ----- 25
Db 4907 HIRVEDIADDCCTLYWMPSSDGGSPITNYIVEKLDLRHSDGKWEKVSFVBNLYTVGG 4966
QY 26 ----- 25
Db 4967 LIKDNRYRFRVRAETQYGVSEPCELADVVAKYQFEVNPQAPATVRDKDSTWAELEMDP 5026
QY 26 ----- 25
Db 5027 PRDGSKIIGYQYVRDITSSGRWINAKMDSLQCHARVTLGNQNGEPEFRILAKNAAGFS 5086
QY 26 ----- 25
Db 5087 KSPSPERCQLSRGPPGPIHVGAKSIGRNHCTITWMAPLEDGSKITGYNVEIREYG 5146
QY 26 ----- 25
Db 5147 STLMTVASDYNRBEFTYVDKLRFNDEYFRVVAIINAAGKIPSLPSCPRIKIQSGSGSRP 5206
QY 26 ----- 25
Db 5207 QIVVPEDTAOPYNNRAVFTCEAVGRPEPTARWLNRGRELPESSRYREASDGYKFTIK 5266
QY 26 ----- 25
Db 5267 EYWDIDAGEYTVESNPGSDTANLVQAPVIEKDPVNTILPSGDLVRLKIYFSGTA 5326
QY 26 ----- 25
Db 5327 PRRHSVLNREEIDMDHPTIRIVERFDHILITIPALSVREAGREYTVSNDGCAITGFW 5386
QY 26 ----- 25
Db 5387 IANTGLPEAPOGLHISNIGPSTATLSWRPPTDGSKITSVYVEKRDLSDENWTVTSN 5446
QY 26 ----- 25
Db 5447 VKDMNYITGLFENNEYEFVSAQNEGIGAPLYSEHPIARLPDPPTSPLEIYQVG 5506
QY 26 ----- 25
Db 5507 GDYVTLNMQRLSDGGRLRGYIVEKQEEHDEWPRCQNPSPNNNVNPLIDGRKRY 5566
QY 26 ----- 25
Db 5567 RVFAVNDAGLSDELHLDQTLFOAGSGGPKIVPSLSDINEVGRCYTFECEISGSRPE 5626
QY 26 ----- 25
Db 5627 YRMFKGKELVDTSKTYILINKGDKQVLLINDLTSDDADEYTCRAITNSGSTRSTANLRK 5686
QY 26 ----- 25
Db 5687 TKPRVFIPIKYHGYEAKGGETIELKIPYKAYPOGEARWTKDGEKIEHNSNFSITTDKF 5746
QY 26 ----- 25
Db 5747 ATLRIASREDYGYRVVENSVDGSIYVNTVADVPEPPRPITENILDEAVILSWK 5806
QY 26 ----- 25
Db 5807 PPALDGSILVNTYTIKREAMGSMSPCAKSRVYTTIEGLRAGKQYEFRIINAKHGOS 5866
QY 26 ----- 25
Db 5867 KPCEPTAVLILPGDEKRRRGRGYDDEQGIYRGKTVSSNVDNYVPDIMKQYPOVEIK 5926
QY 26 ----- 25

```

Db 5927 HDHVIDHDIDIEHELGTGAFGVVHRTERTATGNNFAAKFVMPHESDKETVRKEIQMSYL 5986
QY 26 ----- 25
Db 5987 RHPTLVNLDHAFEDDNEVMYIEFMGSGELFEKVADENHKMSDEAVEYMRQVCKLCHM 6046
QY 26 -----R 26
Db 6047 HENNYVHLDLKPENIMFTTKRSNELKLIDFGLJAHLDPRQSVKVTGTAEFAAPEVAEGK 6106
QY 27 PVGRF 31
    III:
Db 6107 PVGYT 6111
    
```

Search completed: May 1, 2003, 14:49:53
 Job time : 69 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 14:49:32 ; Search time 45 Seconds

(without alignments)
63.277 Million cell updates/sec

Title: SE074-PLUS-73

Perfect score: 156

Sequence: 1 SRHXHSMEXKRTPDINPAMYXXRGIRPVGRFX 33

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.0

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	147	94.2	31	10	US-09-932-161-14
2	147	94.2	31	10	US-09-932-161-15
3	147	94.2	31	12	US-10-044-592-4
4	147	94.2	31	12	US-10-044-592-5
5	147	94.2	70	12	US-10-044-592-90
6	147	94.2	82	12	US-10-044-592-1
7	147	94.2	86	12	US-10-044-592-96
8	147	94.2	87	12	US-10-044-592-92
9	147	94.2	91	12	US-10-044-592-94
10	146	93.6	31	10	US-09-932-161-13
11	146	93.6	31	12	US-10-044-592-39
12	146	93.6	33	12	US-10-044-592-41
13	146	93.6	33	12	US-10-044-592-28
14	146	93.6	98	12	US-10-044-592-38
15	146	93.6	98	12	US-10-044-592-82
16	146	93.6	98	12	US-10-044-592-84
17	146	93.6	98	12	US-10-044-592-86
18	146	93.6	98	12	US-10-044-592-88

20	135	86.5	29	12	US-10-044-592-26	Sequence 26, Appl
21	116	74.4	25	12	US-10-044-592-78	Sequence 78, Appl
22	105	67.3	20	10	US-09-932-161-18	Sequence 18, Appl
23	104	66.7	20	10	US-09-932-161-16	Sequence 16, Appl
24	104	66.7	20	12	US-10-044-592-6	Sequence 17, Appl
25	104	66.7	20	12	US-10-044-592-42	Sequence 6, Appl
26	104	66.7	21	12	US-10-044-592-43	Sequence 43, Appl
27	104	66.7	22	12	US-10-044-592-44	Sequence 44, Appl
28	104	66.7	22	12	US-10-044-592-47	Sequence 27, Appl
29	98	62.8	19	12	US-10-044-592-80	Sequence 80, Appl
30	81	51.9	40	12	US-09-861-289-6	Sequence 6, Appl
31	68	43.6	11877	10	US-09-988-384B-6	Sequence 6, Appl
32	68	43.6	12199	9	US-10-097-354-12	Sequence 12, Appl
33	66	42.3	870	9	US-10-185-050-48	Sequence 48, Appl
34	66	42.3	906	9	US-10-077-130-5	Sequence 2, Appl
35	66	42.3	26926	9	US-09-759-508B-2	Sequence 5, Appl
36	64	41.0	7968	9	US-10-077-130-5	Sequence 4434, Ap
37	64	41.0	2969	9	US-09-738-626-4434	Sequence 3, Appl
38	62	39.7	7257	61	US-10-014-717-3	Sequence 15, Appl
39	61	39.1	1410	9	US-09-966-782A-15	Sequence 8, Appl
40	60	38.5	1410	9	US-10-212-980-8	Sequence 16, Appl
41	59	37.8	371	9	US-09-966-782A-16	Sequence 9, Appl
42	59	37.8	371	9	US-10-212-980-9	Sequence 10447, A
43	59	37.8	372	9	US-09-815-242-10447	
44	59	37.8	940	10		
45	59	37.8				

ALIGNMENTS

```
RESULT 1
US-09-932-161-14
Sequence 14, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Civeili, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Screening and Therapeutic Methods For
FILE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 31
TYPE: PRT
ORGANISM: Rattus
US-09-932-161-14
Query Match          94.2%   Score 147, DB 10; Length 31;
Best Local Similarity 83.9%   Pred. No. 9, 4e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 SRHXHSMEXKRTPDINPAMYXXRGIRPVGRF 31
Db 1 SRAHQHSMETRTPDINPAMYXXRGIRPVGRF 31
RESULT 2
US-09-932-161-15
Sequence 15, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Civeili, Olivier
APPLICANT: Lin, Steven
TITLE OF INVENTION: Screening and Therapeutic Methods For
FILE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
```

```

; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/560,915
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 15
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-932-161-15

```

```

Query Match      94.2%; Score 147; DB 10; Length 31;
Best Local Similarity 83.9%; Pred. No. 9,4e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRAHQSHMETRTPDINPAMYTGIRPVGRF 31

```

```

RESULT 3
US-10-044-592-4
; Sequence 4, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR FILING DATE: 1998-04-27
; PRIOR FILING DATE: 1997-04-28
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO: 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Murine
US-10-044-592-4

```

```

Query Match      94.2%; Score 147; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 9,4e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRAHQSHMETRTPDINPAMYTGIRPVGRF 31

```

```

RESULT 4
US-10-044-592-5
; Sequence 5, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR FILING DATE: 1998-04-27
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:

```

```

; SEQ ID NO: 5
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(31)
; OTHER INFORMATION: antigen
US-10-044-592-5

```

```

Query Match      94.2%; Score 147; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 9,4e-11;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 1 SRAHQSHMETRTPDINPAMYTGIRPVGRF 31

```

```

RESULT 5
US-10-044-592-90
; Sequence 90, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR FILING DATE: 1998-04-27
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO: 90
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Rat
US-10-044-592-90

```

```

Query Match      94.2%; Score 147; DB 12; Length 70;
Best Local Similarity 83.9%; Pred. No. 3e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPVGRF 31
DB 22 SRAHQSHMETRTPDINPAMYTGIRPVGRF 52

```

```

RESULT 6
US-10-044-592-1
; Sequence 1, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:
; APPLICANT: Hinuma, Shuji
; TITLE OF INVENTION: Polypeptides, their Production and Use
; FILE REFERENCE: 2463US2P
; CURRENT APPLICATION NUMBER: US/10/044,592
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/403639
; PRIOR FILING DATE: 1999-25-10
; PRIOR FILING DATE: 1998-04-27
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE:
; SEQ ID NO: 1

```

LENGTH: 82
TYPE: PRT
ORGANISM: Murine
US-10-044-592-1

Query Match 94.28; Score 147; DB 12; Length 82;
Best Local Similarity 83.98; Pred. No. 3.8e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMWYXXRGIRPVGRF 31
DB 21 SRAHQHSMETRTPDINPAMWYTGRCIRPVGRF 51

RESULT 7

US-10-044-592-96
Sequence 96, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 96
LENGTH: 86
TYPE: PRT
ORGANISM: mammalian
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1224)..(1243)
OTHER INFORMATION: Bracket region depicted in FIG 39.
US-10-044-592-96

Query Match 94.28; Score 147; DB 12; Length 86;
Best Local Similarity 83.98; Pred. No. 4.1e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 SRXHXSMEXRTPDINPAMWYXXRGIRPVGRF 31
DB 22 SRAHQHSMETRTPDINPAMWYTGRCIRPVGRF 52

RESULT 8

US-10-044-592-92
Sequence 92, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 92
LENGTH: 87

TYPE: PRT
ORGANISM: Homo sapiens
US-10-044-592-92

Query Match 94.28; Score 147; DB 12; Length 87;
Best Local Similarity 83.98; Pred. No. 4.1e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMWYXXRGIRPVGRF 31
DB 23 SRAHQHSMETRTPDINPAMWYASRGIRPVGRF 53

RESULT 9

US-10-044-592-94
Sequence 94, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Fukusumi, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT APPLICATION NUMBER: US/10/044,592
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: PCT/JP98/01923
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 94
LENGTH: 91
TYPE: PRT
ORGANISM: Mammalian
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(31)
OTHER INFORMATION: primer
NAME/KEY: misc_feature
LOCATION: (925)..(955)
OTHER INFORMATION: primer
NAME/KEY: misc_feature
LOCATION: (1)..(955)
OTHER INFORMATION: Insert fragment of pmGB3 as depicted in FIG 36 and 37.
US-10-044-592-94

Query Match 94.28; Score 147; DB 12; Length 91;
Best Local Similarity 83.98; Pred. No. 4.4e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMWYXXRGIRPVGRF 31
DB 22 SRAHQHSMETRTPDINPAMWYTGRCIRPVGRF 52

RESULT 10

US-09-932-161-13
Sequence 13, Application US/09932161
Patent No. US20020037533A1
GENERAL INFORMATION:
APPLICANT: Lin, Steven
APPLICANT: Cleveli, Olivier
TITLE OF INVENTION: Screening and Therapeutic Methods for
TITLE OF INVENTION: Promoting Wakefulness and Sleep
FILE REFERENCE: P-UC 4679
CURRENT APPLICATION NUMBER: US/09/932,161
PRIOR FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/560,915
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: fastseq for Windows Version 4.0

Thu May 1 15:54:07 2003

seq74-plus-73.rapb

Page 4

SEQ ID NO 13
LENGTH: 31
TYPE: PRT
ORGANISM: Bos taurus
US-09-932-161-13

Query Match 93.6%; Score 146; DB 10; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPYGRF 31
DB 1 SRAHQSMETRTPDINPAMYAGRGIRPYGRF 31

RESULT 11
US-10-044-592-39
Sequence 39, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT FILING DATE: 2002-01-10
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 39
LENGTH: 31
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-39

Query Match 93.6%; Score 146; DB 12; Length 31;
Best Local Similarity 83.9%; Pred. No. 1.2e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPYGRF 31
DB 1 SRAHQSMETRTPDINPAMYAGRGIRPYGRF 31

RESULT 12
US-10-044-592-40
Sequence 40, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT FILING DATE: 2002-01-10
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 40
LENGTH: 32
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-40

Query Match 93.6%; Score 146; DB 12; Length 32;
Best Local Similarity 83.9%; Pred. No. 1.3e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPYGRF 31
DB 1 SRAHQSMETRTPDINPAMYAGRGIRPYGRF 31

RESULT 13
US-10-044-592-41
Sequence 41, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT FILING DATE: 2002-01-10
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 41
LENGTH: 33
TYPE: PRT
ORGANISM: Bovine
US-10-044-592-41

Query Match 93.6%; Score 146; DB 12; Length 33;
Best Local Similarity 83.9%; Pred. No. 1.3e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 SRXHXSMEXRTPDINPAMYXXRGIRPYGRF 31
DB 1 SRAHQSMETRTPDINPAMYAGRGIRPYGRF 31

RESULT 14
US-10-044-592-28
Sequence 28, Application US/10044592
Patent No. US20020143152A1
GENERAL INFORMATION:
APPLICANT: Hinuma, Shoji
TITLE OF INVENTION: Polypeptides, their Production and Use
FILE REFERENCE: 2463US2P
CURRENT FILING DATE: 2002-01-10
PRIOR FILING DATE: 1999-25-10
PRIOR APPLICATION NUMBER: US 09/403639
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: JP 9-109974
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 96
SOFTWARE:
SEQ ID NO 28
LENGTH: 98
TYPE: PRT
ORGANISM: Murine
US-10-044-592-28

Query Match 93.6%; Score 146; DB 12; Length 98;
Best Local Similarity 83.9%; Pred. No. 6.4e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

seq74-plus-73.rabb

Page 5

RESULT 15
US-10-044

US-10-044-592-38
; Sequence 38, Application US/10044592
; Patent No. US20020143152A1
; GENERAL INFORMATION:

```
; APPLICANT: Hinuma, Shuji
;
; APPLICANT: Fukusumi, Shoji
```

;; TITLE OF INVENTION: Polypeptides, their Production and Use
;; FILE REFERENCE: 24637US2P

CURRENT APPLICATION NUMBER: US/10/044,592
CURRENT FILING DATE: 2003-03-10

COMMENT FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 09/403639

; PRIOR FILING DATE: 1999-25-10
; PRIOR APPLICATION NUMBER: PCT/JP98/01923

; PRIOR FILING DATE: 1998-04-27
 ; PRIOR APPLICATION NUMBER: JP 9-10997A

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

SOFTWARE: 3000
; SOFTWARE: 3000
; SOFTWARE: 3000

; SEX: ID NO 38
; LENGTH: 98; TYPE: PRT
; ORGANISM:

US-10-044-592-38

Query Match	Score	DB	Length
Best Local Similarity:	93.68;	146;	12;
	97.00;	12;	98;

best local similarity 83.9%; Pred. No. 6.4e-10;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0.

QY 1 SRXHXHSMEXRTPDINPAWYXXRGIRPVGRF 31

```
Db      |||||
23 SRAHSHMEIRTPDINPAWYAGRGITPVGPE 53
```

Search completed: May 1, 2003, 14:54:37
Job time : 45 secs

